Introduction to Artificial Intelligence

丁尧相
浙江大学

Fall & Winter 2022
Week 6

Announcements

- The Problem Set 1 will be returned next Monday.
- About Lab Project 1.

Knowledge Reasoning: II

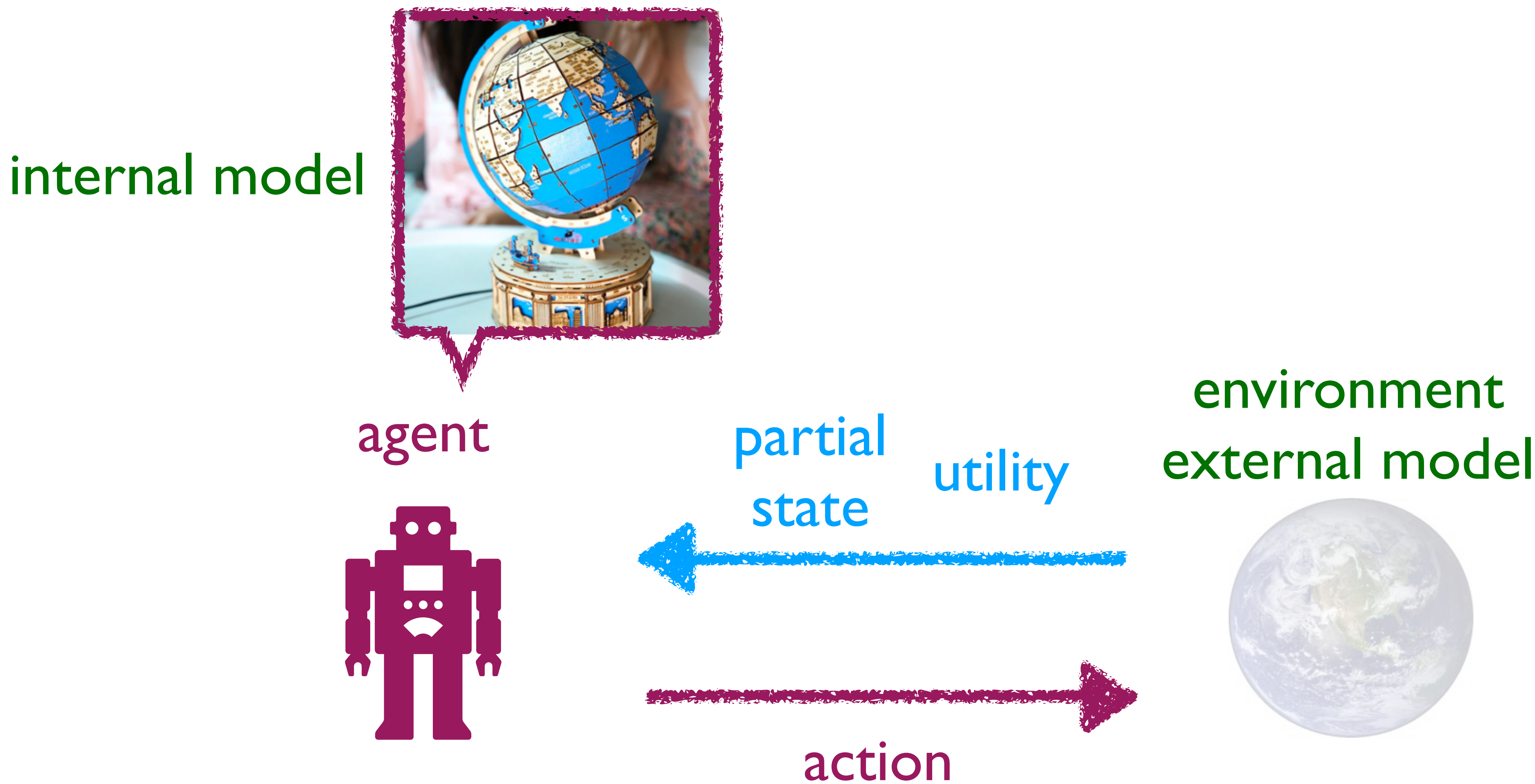
- Probabilistic reasoning
- The structure of probability distributions
- Bayes net: Representation
- Bayes net: Conditional independence
- Bayes net: Exact inference
- Take-home messages

Knowledge Reasoning: II

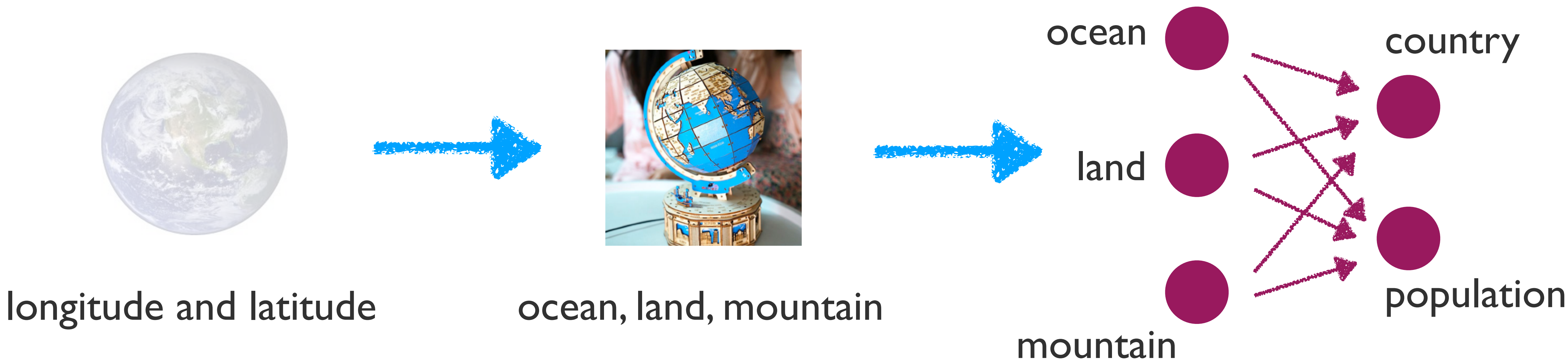
- Probabilistic reasoning
- The structure of probability distributions
- Bayes net: Representation
- Bayes net: Conditional independence
- Bayes net: Exact inference
- Take-home messages

Internal vs. External Model

Since the agent cannot fully know the external model,
it should build an internal model itself for decision making.



Knowledge in AI Systems

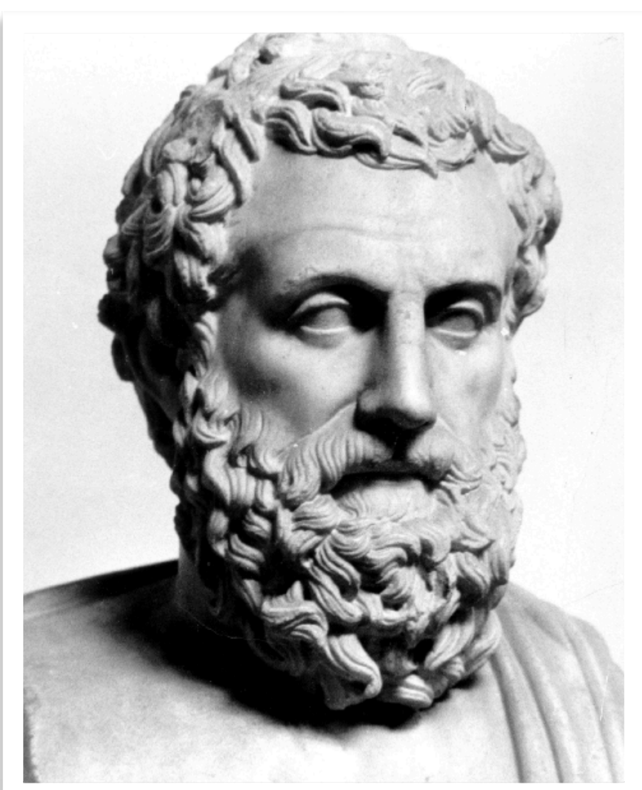
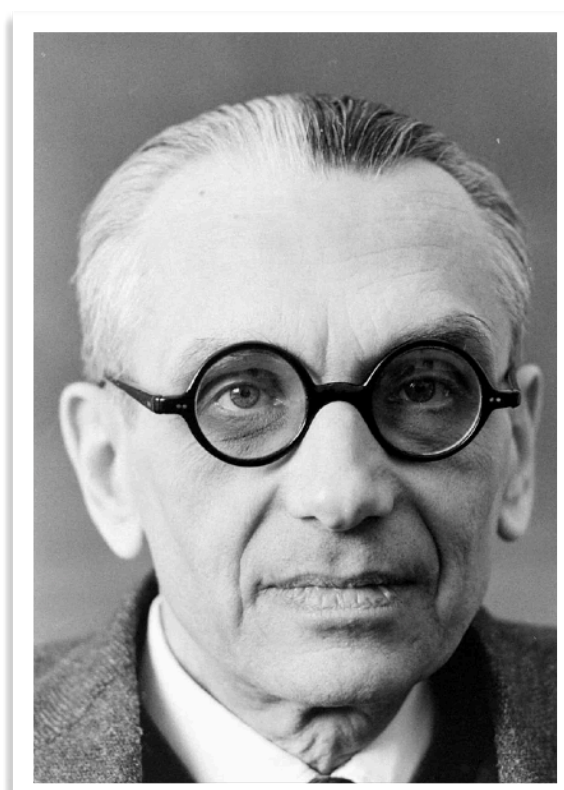
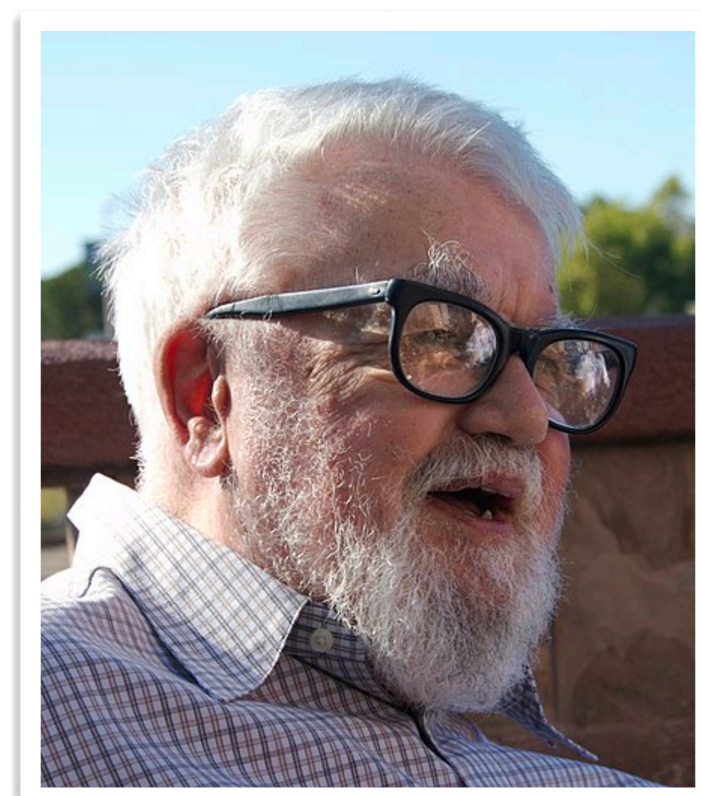


- Turn primitive external states into meaningful internal states.
- Reason about most useful states for decision making.
- Capture internal relationships among factors of decision making.

These reasoning rules are called knowledges in an AI system.

Logic Reasoning Systems

- Handling decision problems (true/false arguments).
- Handling discrete and (not exactly) deterministic world.

From Deterministic to Probabilistic Reasoning

- Modeling human knowledge, e.g. common sense:
  - Uncertainty in reasoning
  - Fast learning and processing
- We will introduce two foundational tools:
  - Bayes nets
  - Causal reasoning

Judea Pearl

Knowledge Reasoning: II

- Probabilistic reasoning
- The structure of probability distributions
- Bayes net: Representation
- Bayes net: Conditional independence
- Bayes net: Exact inference
- Take-home messages

Probability Distributions

- Joint distribution: $P(X_1, X_2, ..., X_n)$
- Marginal distribution: $P(X_1, X_2)$
- Conditional distribution: $P(X_3|X_1, X_2)$
- The meaning of probabilities?

Probability Distributions

- Joint distribution: $P(X_1, X_2, ..., X_n)$
- Marginal distribution: $P(X_1, X_2)$
- Conditional distribution: $P(X_3|X_1, X_2)$
- The meaning of probabilities?

In frequentist understanding,
the probability is the rule behind repeated random trials.

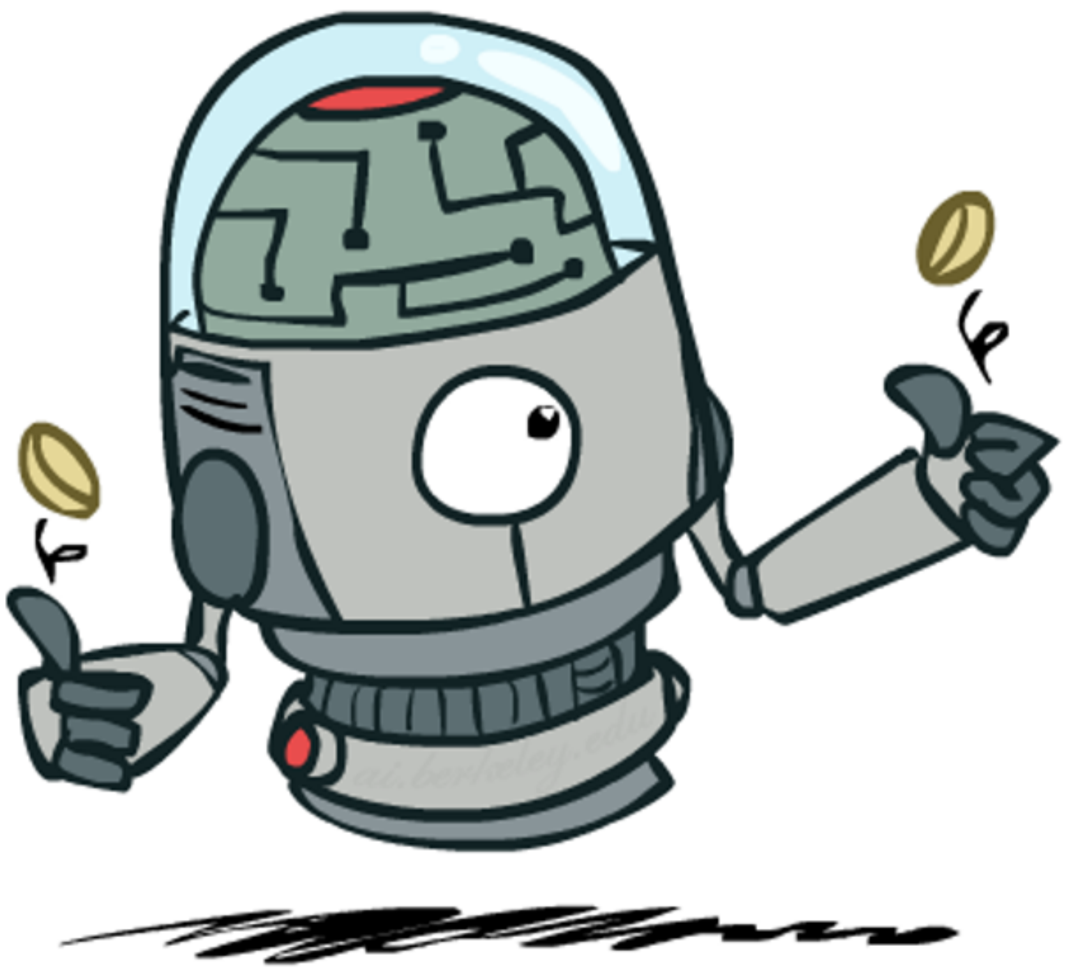
Probability Distributions

- Joint distribution: $P(X_1, X_2, ..., X_n)$
- Marginal distribution: $P(X_1, X_2)$
- Conditional distribution: $P(X_3|X_1, X_2)$
- The meaning of probabilities?

In frequentist understanding,
the probability is the rule behind repeated random trials.

In Bayesian understanding,
the probability is the measure of uncertainty for real-world events.
Not necessary about repeated events! e.g. the probability tomorrow rains.

Independence



Slide courtesy: Dan Klein & Pieter Abbeel

Independence (cont.)

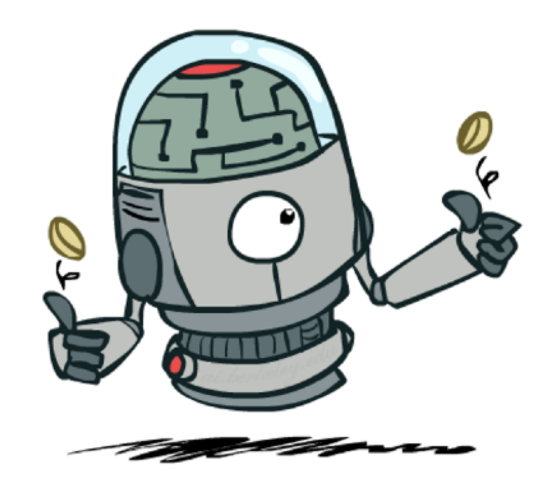
- Two random variables are independent if:

$$\forall x, y : P(x, y) = P(x)P(y)$$

- The joint distribution factors into the product of two simpler distributions.

- Another form: $\forall x, y : P(x|y) = P(x)$

- Notation: $X \perp\!\!\!\perp Y$



Independence is a powerful tool to find simplified structures of probability distributions.

Slide courtesy: Stuart Russell & Sergey Levine

Example

- Whether independence holds?

$P(T)$

T	P
hot	0.5
cold	0.5

$P(W)$

W	P
sun	0.6
rain	0.4

$P_1(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$P_2(T, W)$

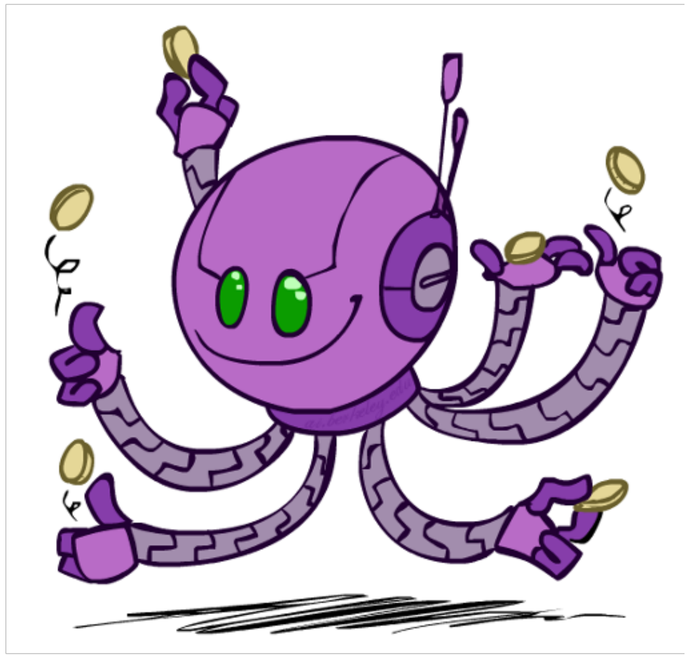
T	W	P
hot	sun	0.3
hot	rain	0.2
cold	sun	0.3
cold	rain	0.2

Slide courtesy: Dan Klein & Pieter Abbeel

Mutual Independence vs. Pairwise Independence

- For a set of random variables,
  - Mutual independence: any subset of r. v. are independent.
  - Pairwise independence: any pair of r.v. are independent.

pairwise independence is a weaker property than mutual independence!



Slide courtesy: Dan Klein & Pieter Abbeel

Conditional Independence

- X_1 : whether sleep yesterday
  X_2 : whether study hard today
  X_3 : whether final exam pass

- Conditional independence:

$$P(X_3|X_1, X_2) = P(X_3|X_2)$$

$$P(X_1|X_2, X_3) = P(X_1|X_2)$$

$$P(X_1, X_3|X_2) = P(X_1|X_2)P(X_3|X_2)$$

Independence holds only when we know the value of the condition.

The Chain Rule

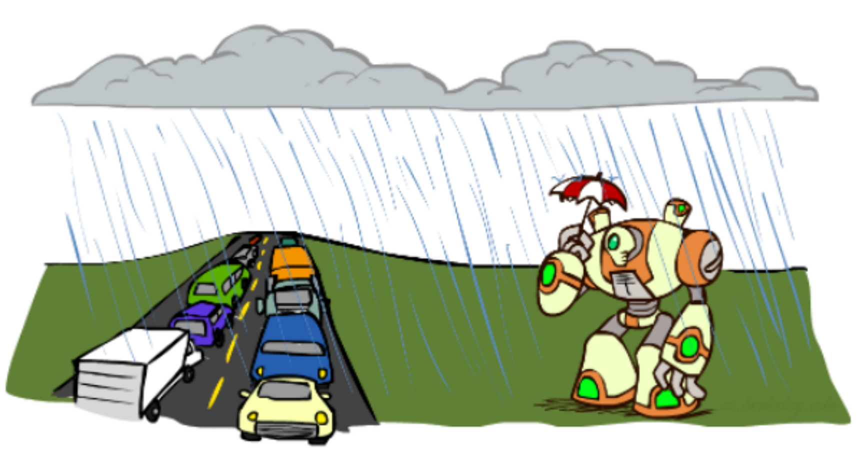
- **Chain rule:** $P(X_1, X_2, \ldots X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots$

- **Trivial decomposition:**

$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) =$$
$$P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain}, \text{Traffic})$$

- **With assumption of conditional independence:**

$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) =$$
$$P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain})$$



Slide courtesy: Stuart Russell & Sergey Levine

Basic Rules

- Given the joint distribution $P(X, Y)$
- Sum rule (marginalization):

$$P(X) = \sum_n P(X, Y = n)$$

- Product rule (factorization):

$$P(X, Y) = P(X)P(Y|X)$$

- Bayes theorem:

$$P(X)P(Y|X) = P(Y)P(X|Y)$$

Hold for any joint distribution, generalize to a set of r.v.
Basic tools in Bayesian inference.

Knowledge Reasoning: II

- Probabilistic reasoning
- The structure of probability distributions
- Bayes net: Representation
- Bayes net: Conditional independence
- Bayes net: Exact inference
- Take-home messages

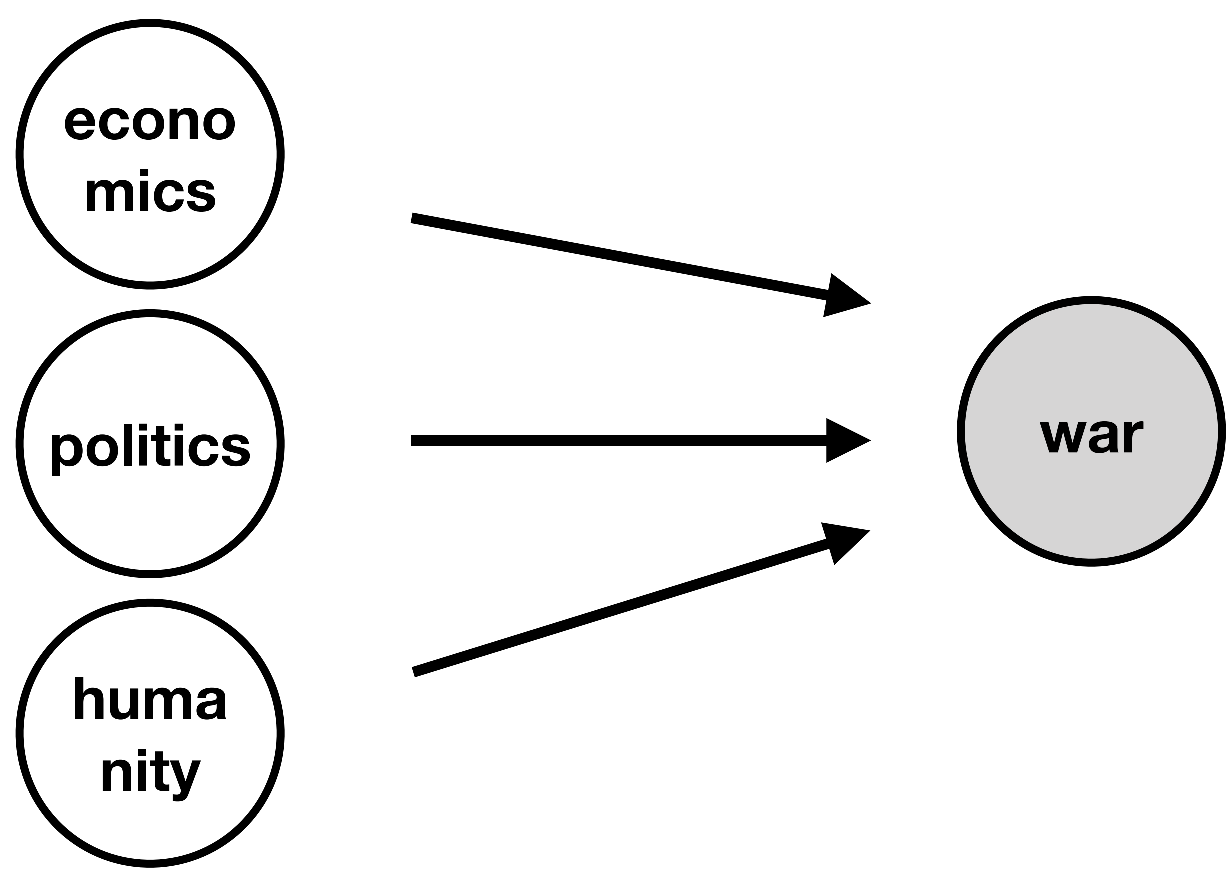
Basic Tasks in Probabilistic Reasoning

- In probabilistic reasoning, we try to model the joint distribution of a set of random variables $P(X_1, X_2, ..., X_n)$ and do:
  - Inference: answering queries about the marginal distributions.
  - Conditional independence test: decide the conditional independence of a subset of random variables.
  - Learning: obtain the structure of the joint distribution.

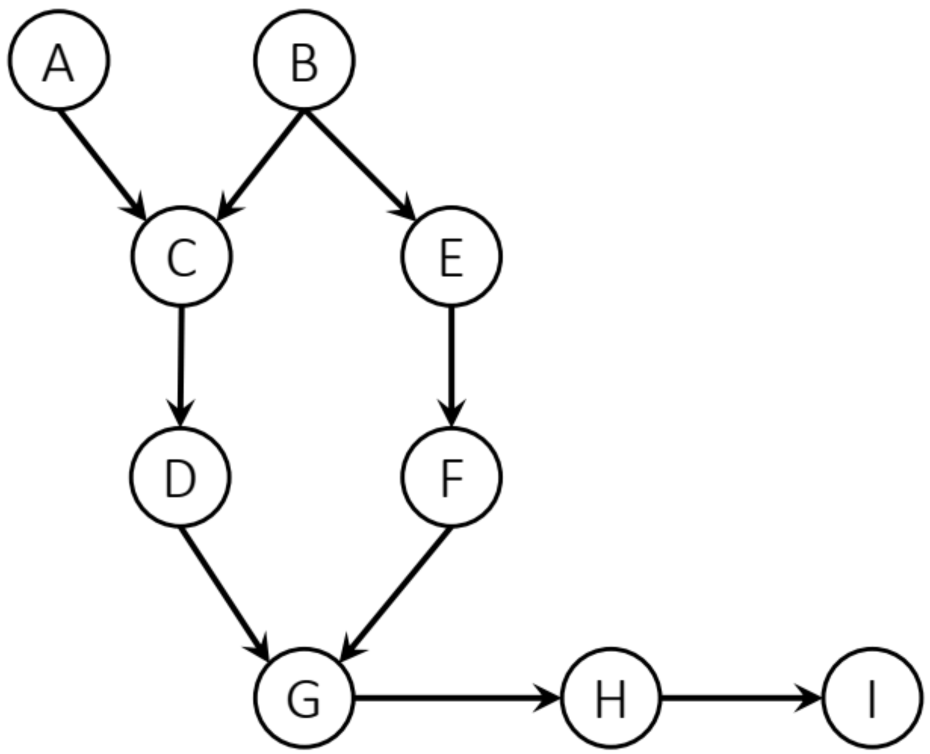
Inference is to reasoning about the value of the variables.
The independence test and learning are to understand relationship among variables.

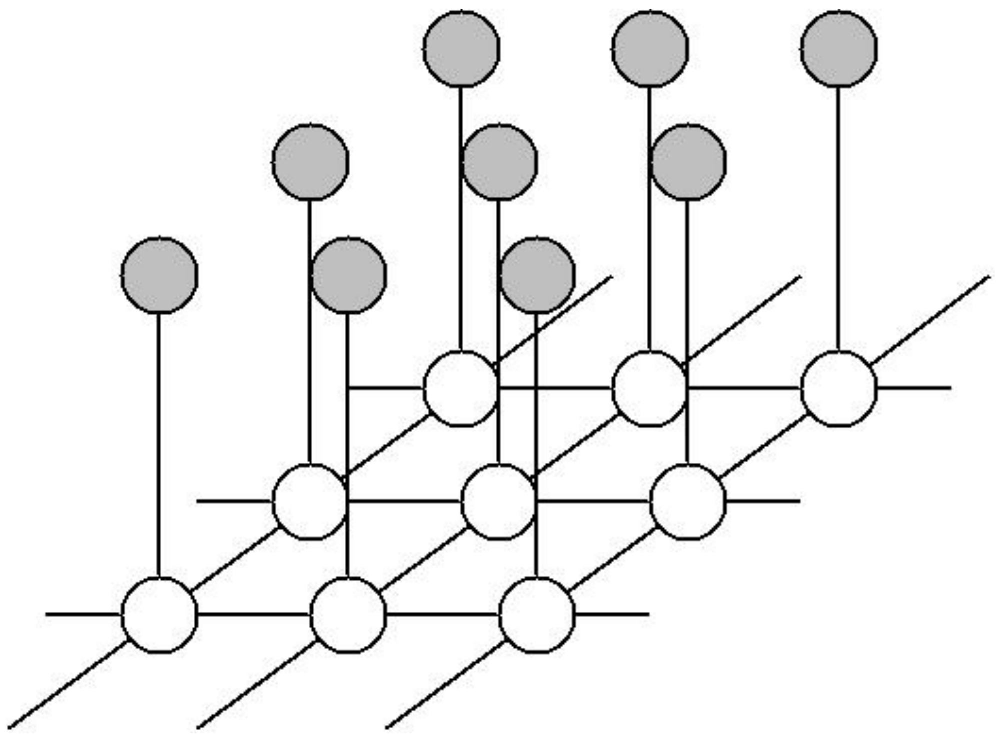
Graphical Models

- Graphical models represent the joint distribution over a set of random variables with directed or undirected graphs.
    - nodes: random variables (can be hidden or observable)
    - edges: the interaction between a pair of r. v.



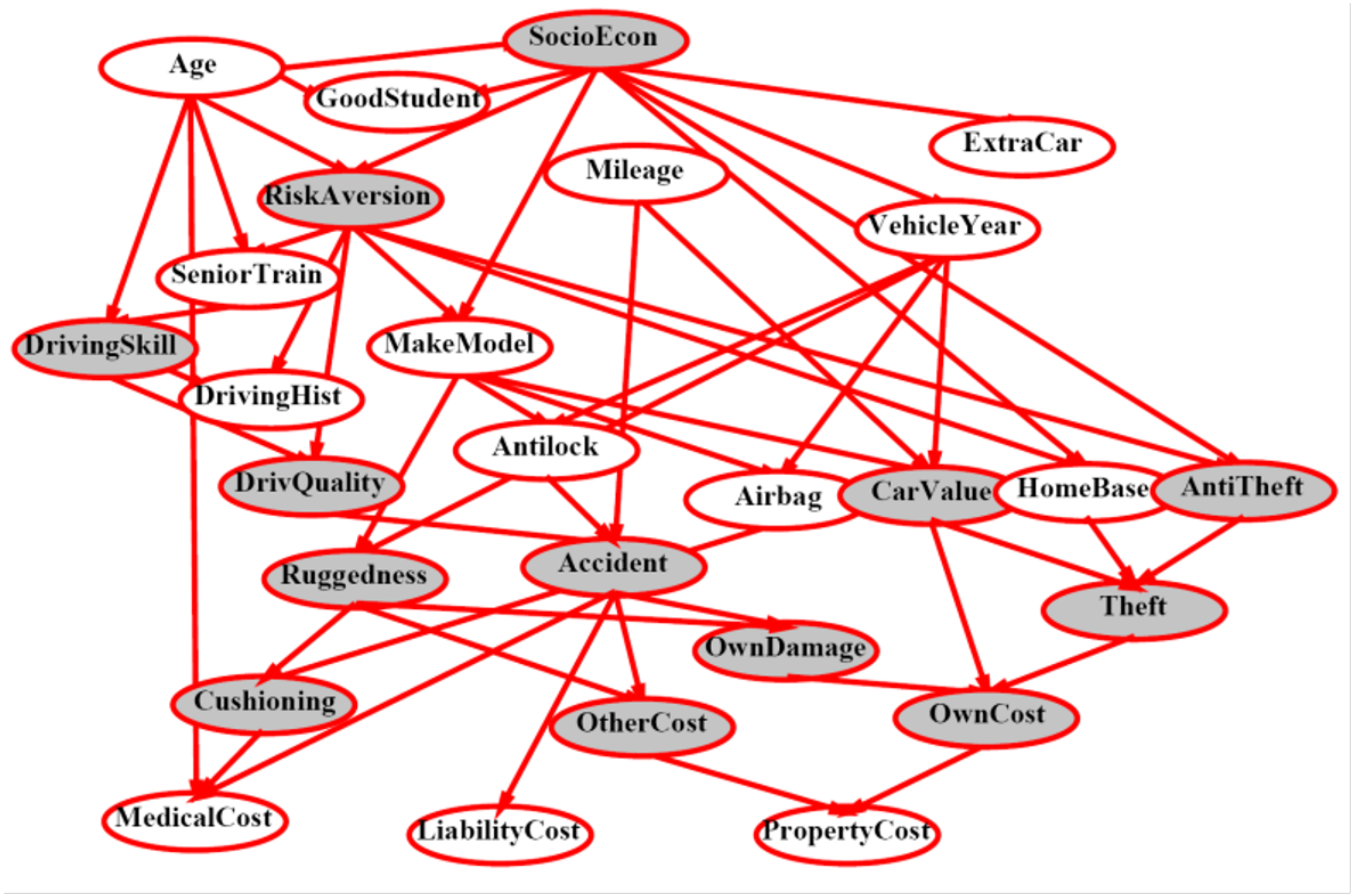
Bayesian Networks & Markov Random Fields



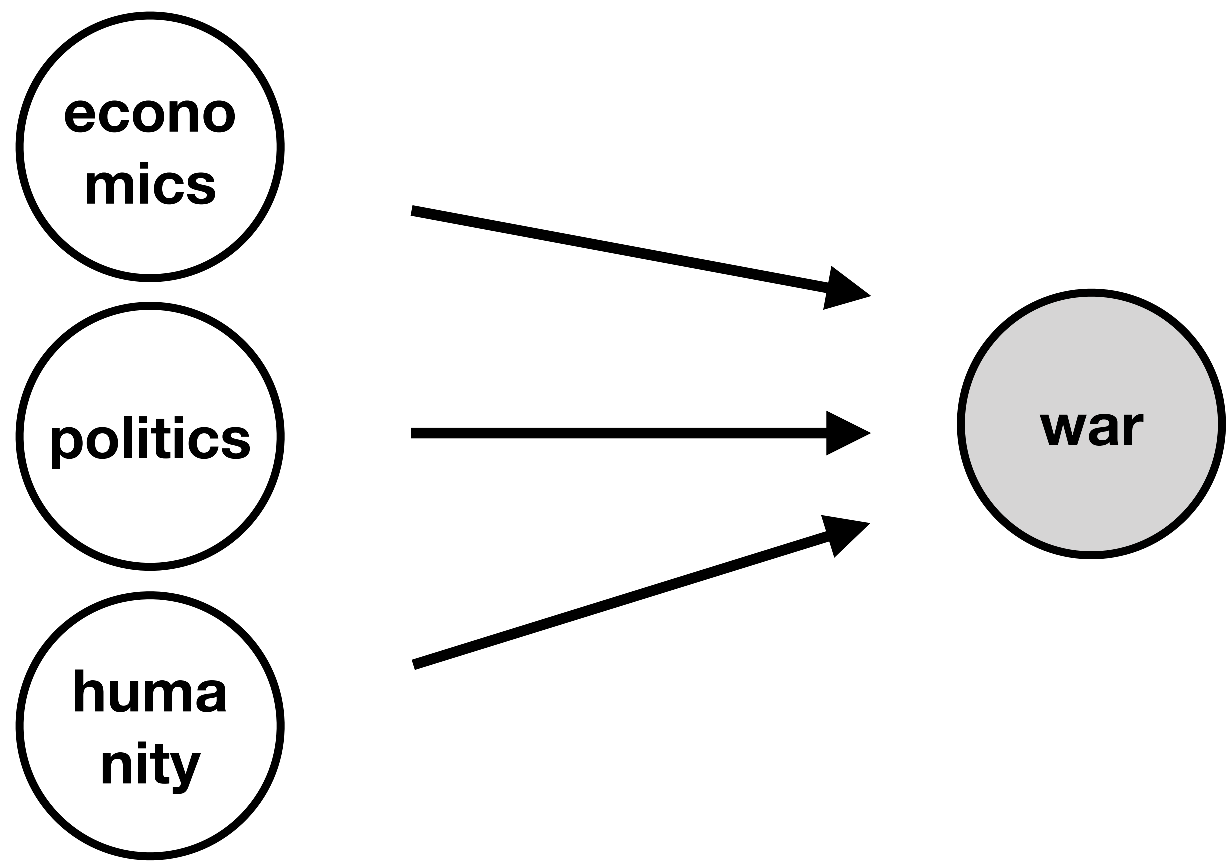


Equivalent representation power!
We will focus on Bayes nets in this course.

Bayes Nets


SocioEcon
Age
GoodStudent
ExtraCar
Mileage
RiskAversion
VehicleYear
SeniorTrain
DrivingSkill
MakeModel
DrivingHist
Antilock
DrivQuality
Airbag
CarValue
HomeBase
AntiTheft
Ruggedness
Accident
Theft
OwnDamage
Cushioning
OtherCost
OwnCost
MedicalCost
LiabilityCost
PropertyCost

Bayes Nets

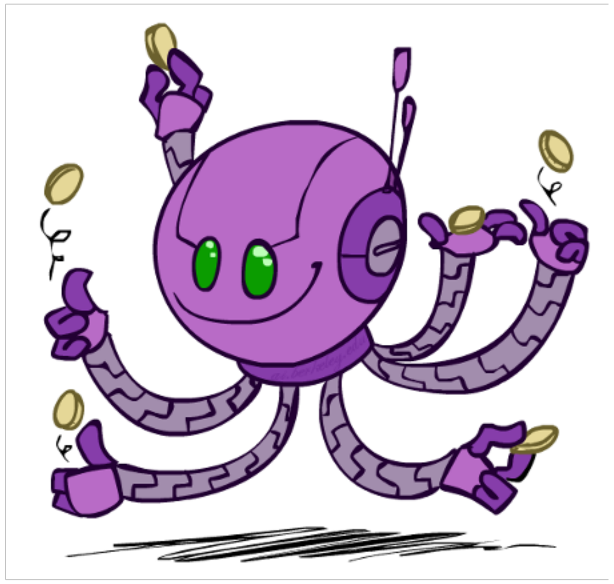


- Bayes nets are acyclic directed graphs representing the joint distribution of random variables.

- The nodes represent random variables, the edges represent the relationship among r. v. (e.g. hidden to observable r. v.)

Bayes Nets: Example

- N independent coin flips

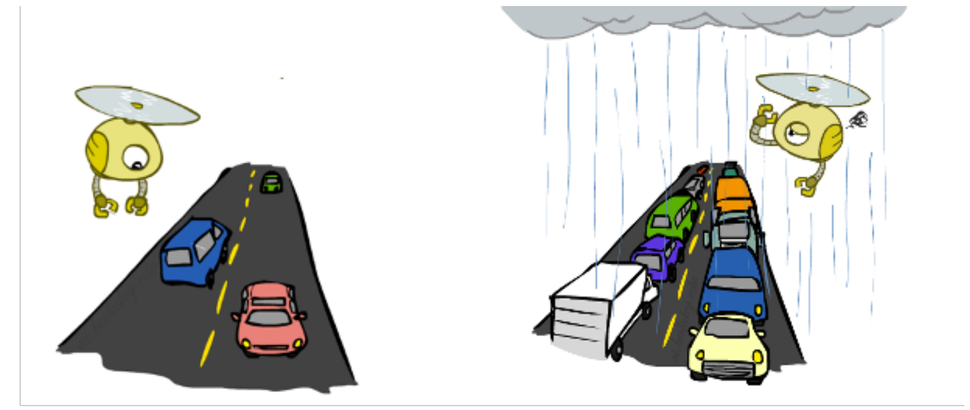


X_1 X_2 ... X_n

- No interactions between variables: absolute independence

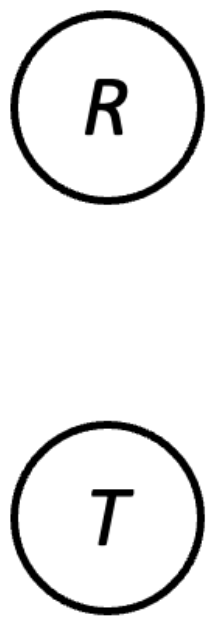
Slide courtesy: Stuart Russell & Sergey Levine

Bayes Nets: Example

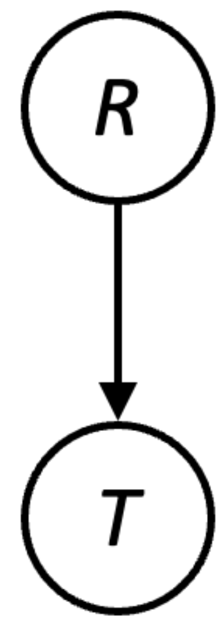


- Variables:
  - R: It rains
  - T: There is traffic

- Model 1: independence



- Model 2: rain causes traffic

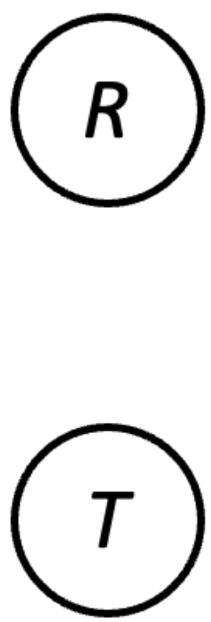


- Why is an agent using model 2 better?

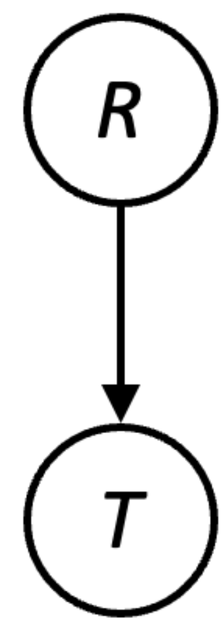
Bayes Nets: Example

- Variables:
  - R: It rains
  - T: There is traffic

- Model 1: independence



- Model 2: rain causes traffic

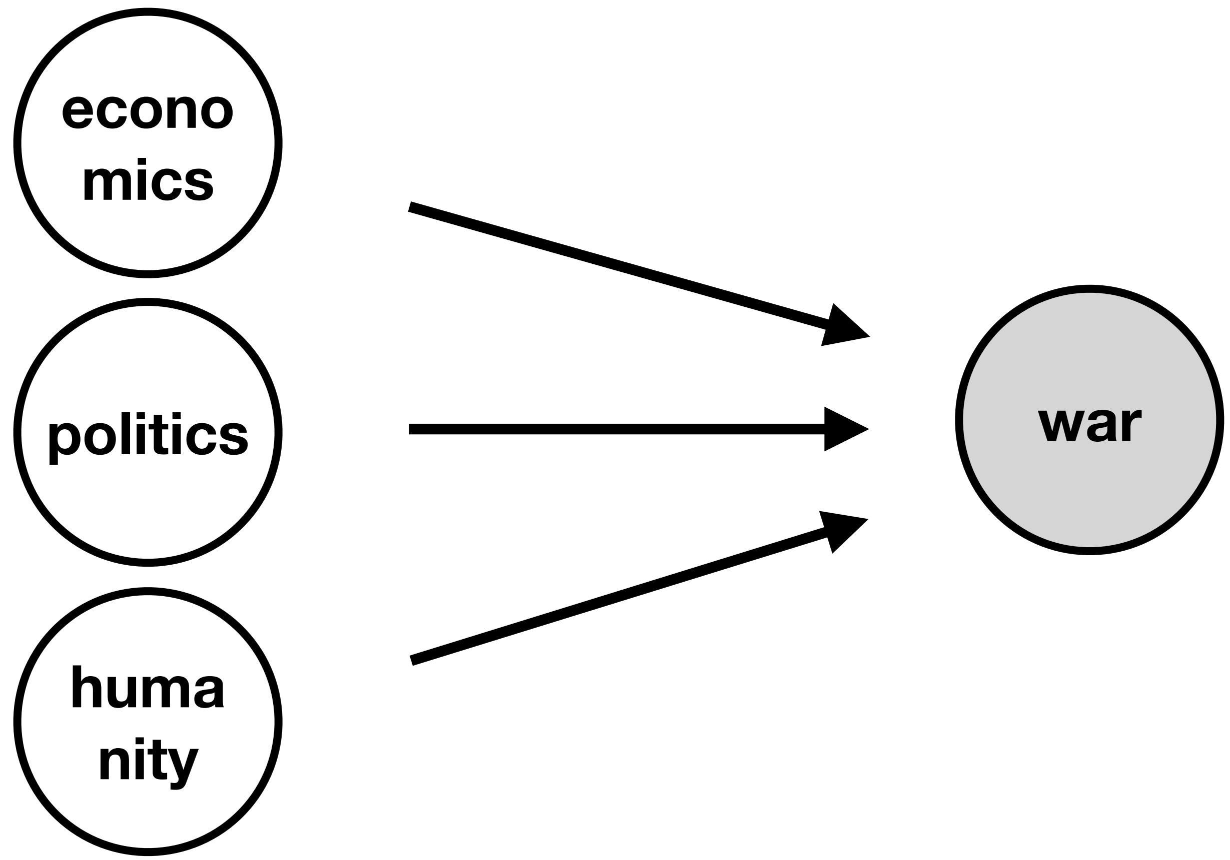


- Why is an agent using model 2 better?

Caution: Bayes net does NOT truly encode causal relationships.
The edge direction only implies the way of factorization!

Slide courtesy: Stuart Russell & Sergey Levine

Bayes Nets: Semantics

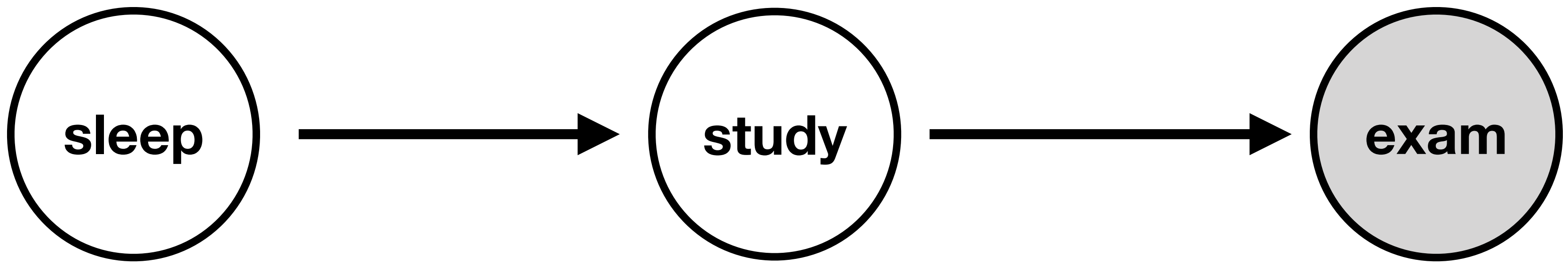


- Bayes nets encode the joint distribution:

$$P(war, eco, pol, hum) = P(war|eco, pol, hum)P(eco)P(pol)P(hum)$$

Factored based on the chain rule

Bayes Nets: Semantics



$$P(sleep, study, exam) = P(exam|study, sleep)P(study|sleep)P(sleep)$$

$$\Downarrow$$

$$P(sleep, study, exam) = P(exam|study)P(study|sleep)P(sleep)$$

Bayes nets represent the factored (conditional independence) information

Knowledge Reasoning: II

- Probabilistic reasoning
- The structure of probability distributions
- Bayes net: Representation
- Bayes net: Conditional independence
- Bayes net: Exact inference
- Take-home messages

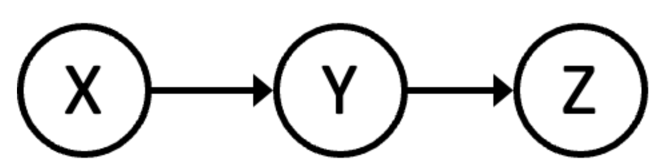
Basic Tasks in Probabilistic Reasoning

- In probabilistic reasoning, we try to model the joint distribution of a set of random variables $P(X_1, X_2, ..., X_n)$ and do:
  - Inference: answering queries about the marginal distributions.
  - Conditional independence test: decide the conditional independence of a subset of random variables.
  - Learning: obtain the structure of the joint distribution.

Conditional Independence

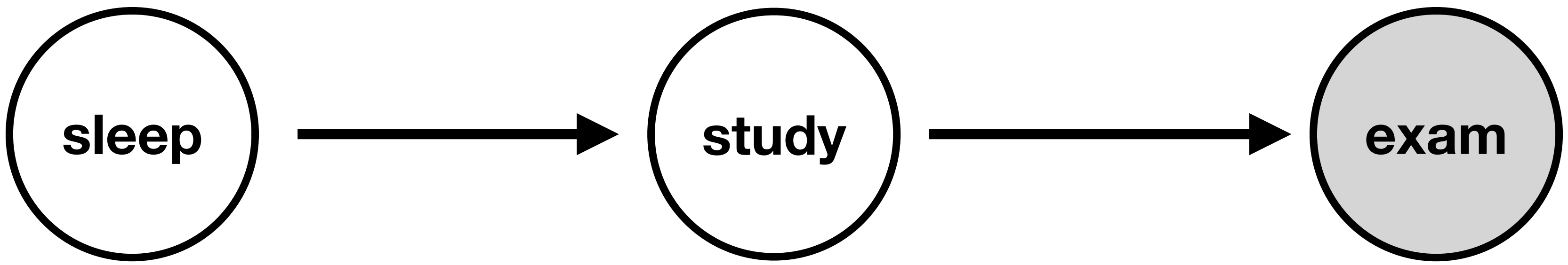
- Conditional independence is the most important structural information in Bayes nets.
  - The more conditional independence properties, the simpler the Bayes net is.

- A fundamental question to answer given a Bayes net:
Are two nodes conditionally independent given other nodes?
  - Lie in the heart of many inference algorithms in BN.



Is X independent of Z?
Is X independent of Z given Y?

Conditional Independence

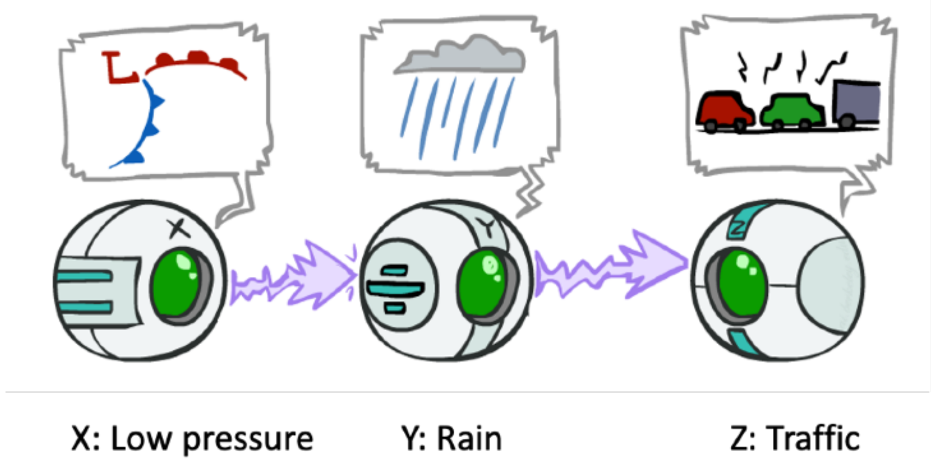


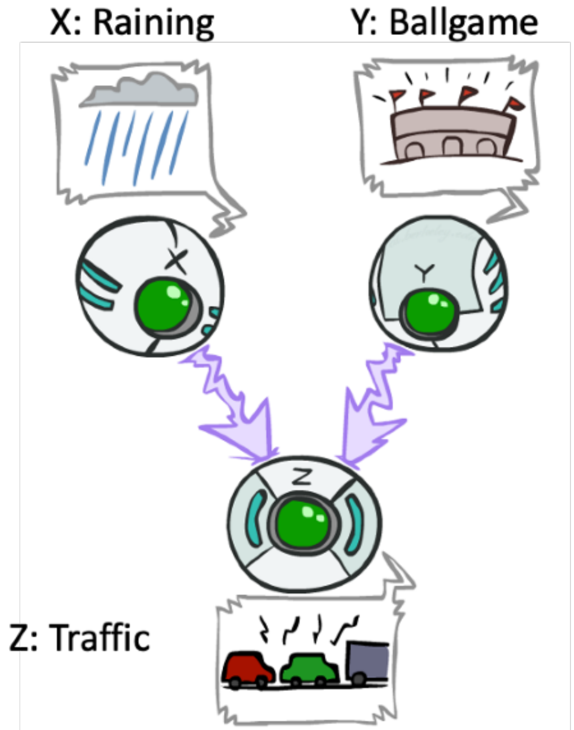
- The simplest case: a chain graph.
- Intuitively, the intermediate variables “blocks” the chain.

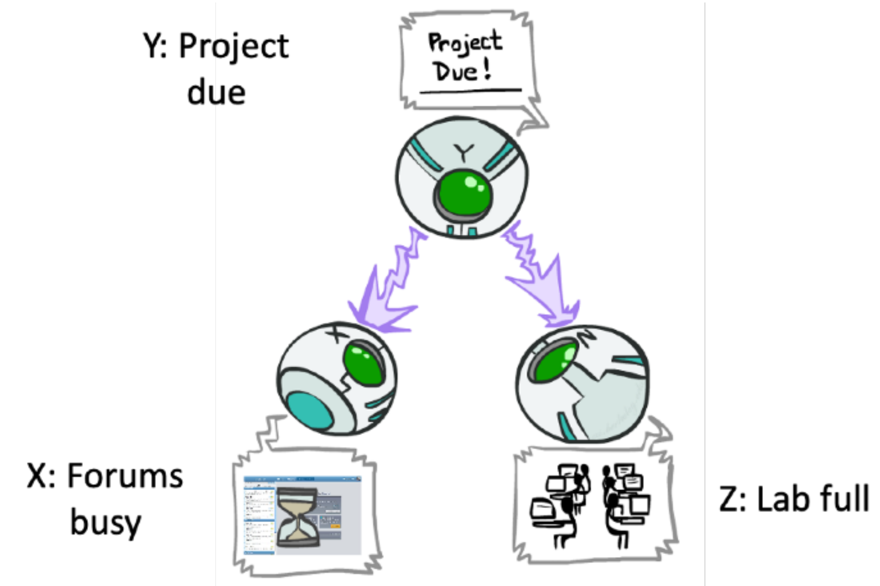
In most situations, the graph does not has such simple structure.
We should makes use of more delicate independence-test rules.

D-Separation

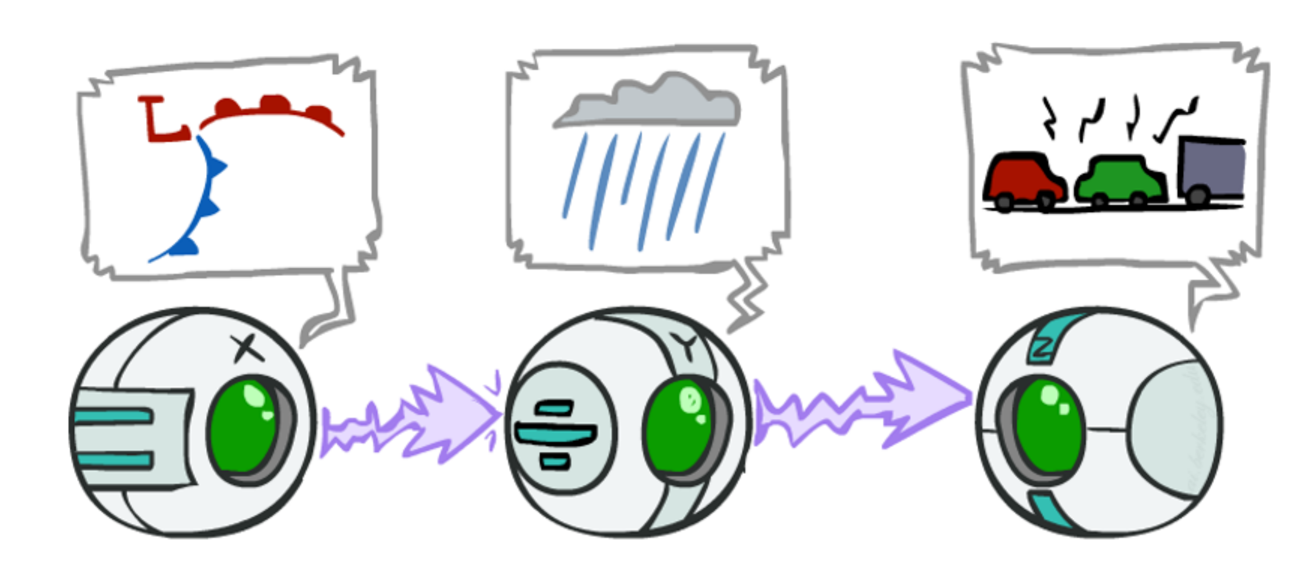
- Idea: break the graph piece by piece: studying triplet relationships



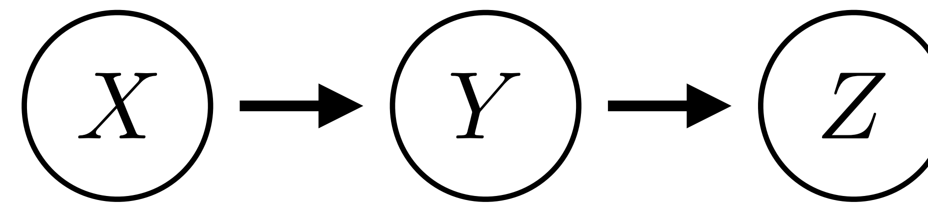




The Triplet Chain



X: Low pressure Y: Rain Z: Traffic



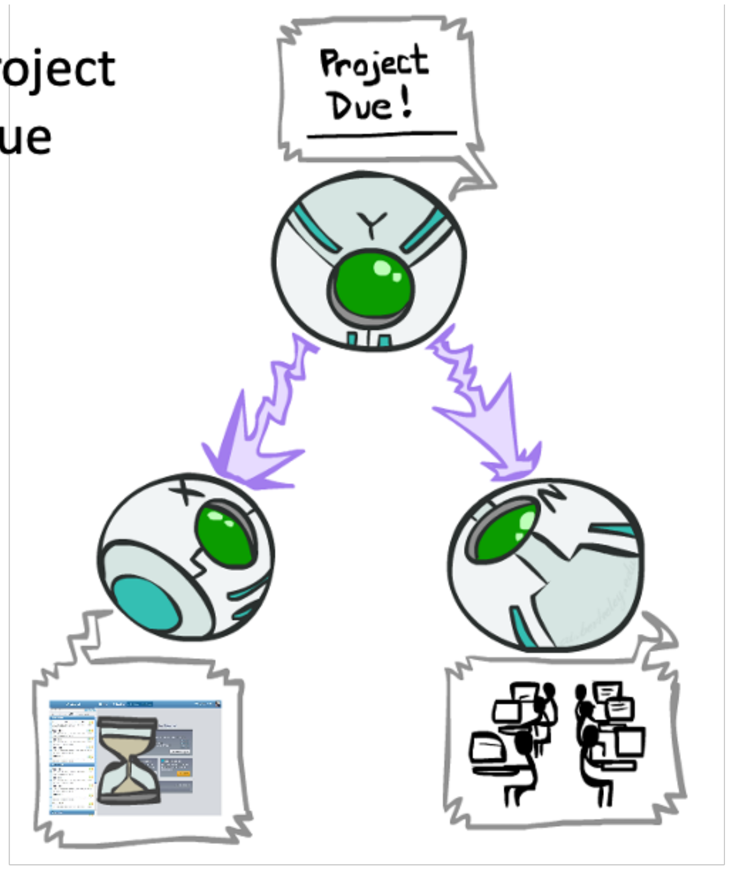
Y "blocks" the information from X to Z

$$P(x, y, z) = P(x)P(y|x)P(z|y)$$

We can prove that $p(Z|Y) = p(Z|X, Y)$
Indicating $X \perp\!\!\!\perp Z|Y$

The Common Cause

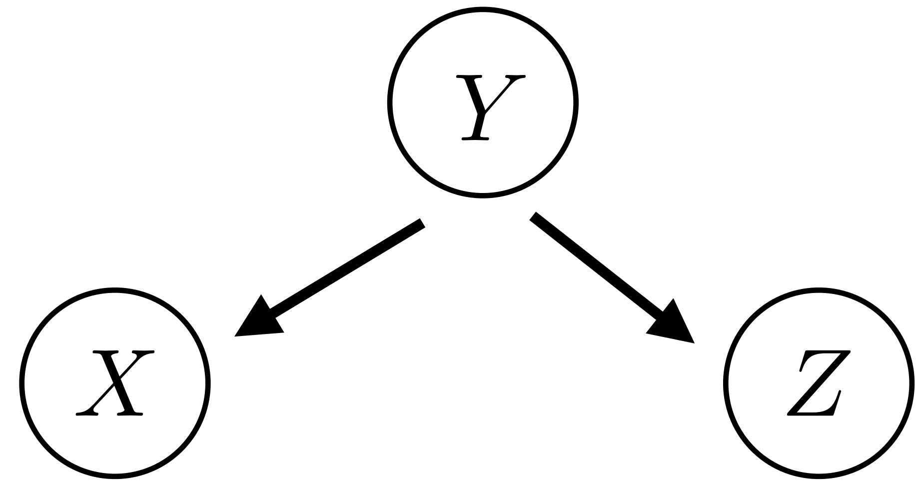
Y: Project due



X: Forums busy

Z: Lab full

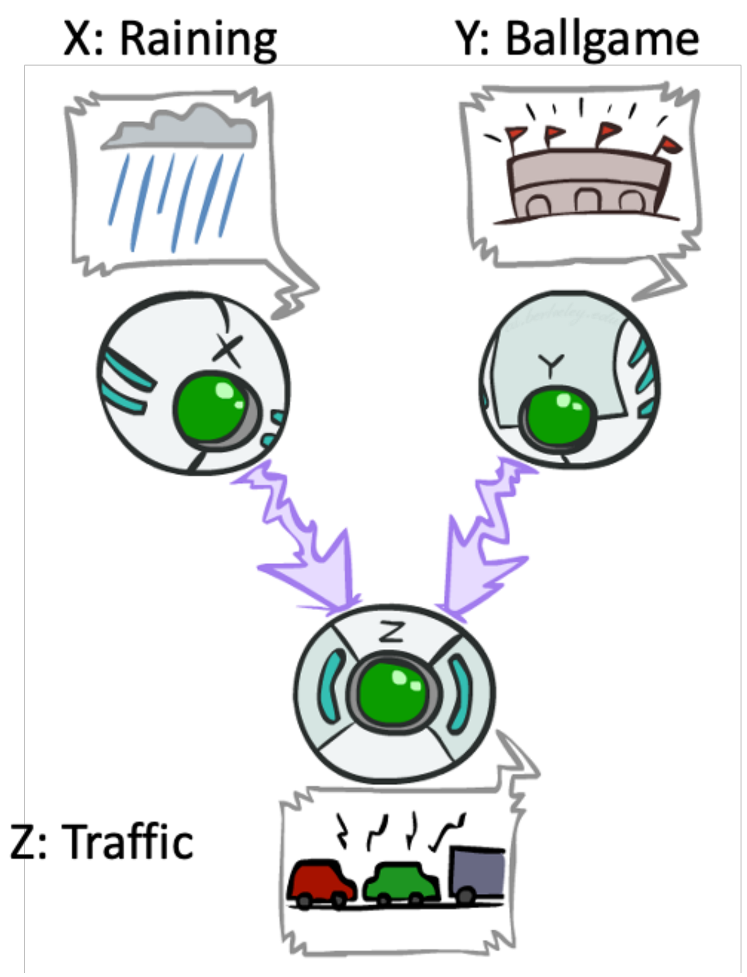
$$P(x, y, z) = P(y)P(x|y)P(z|y)$$

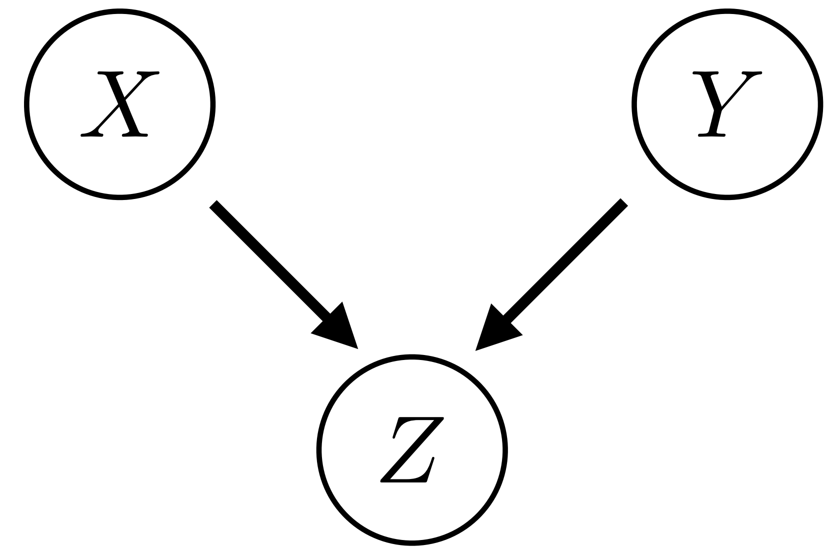


Know the common cause is already enough.

We can prove that $p(Z|Y) = p(Z|X, Y)$
Indicating $X \perp Z|Y$

The Common Effect (V-Structure)



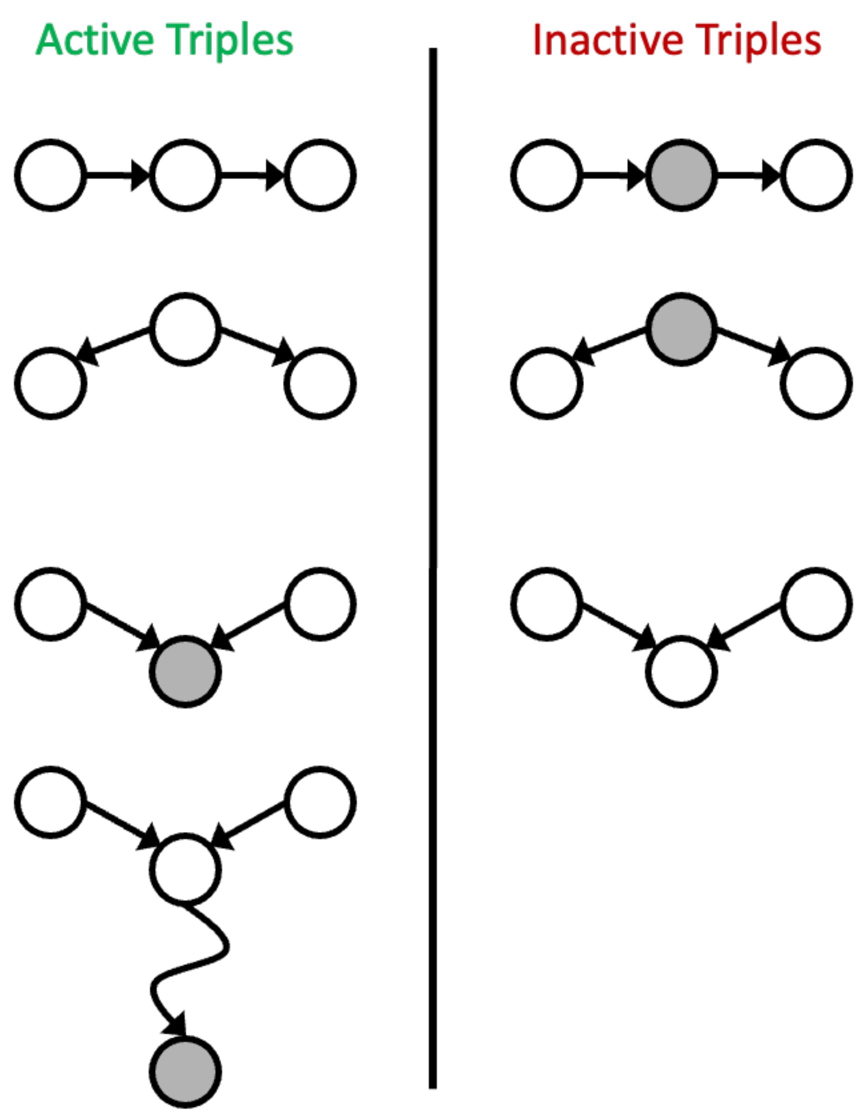


Knowing Z connects X and Y

Even when $X \perp\!\!\!\perp Y$, knowing Z makes X and Y not conditionally independent!

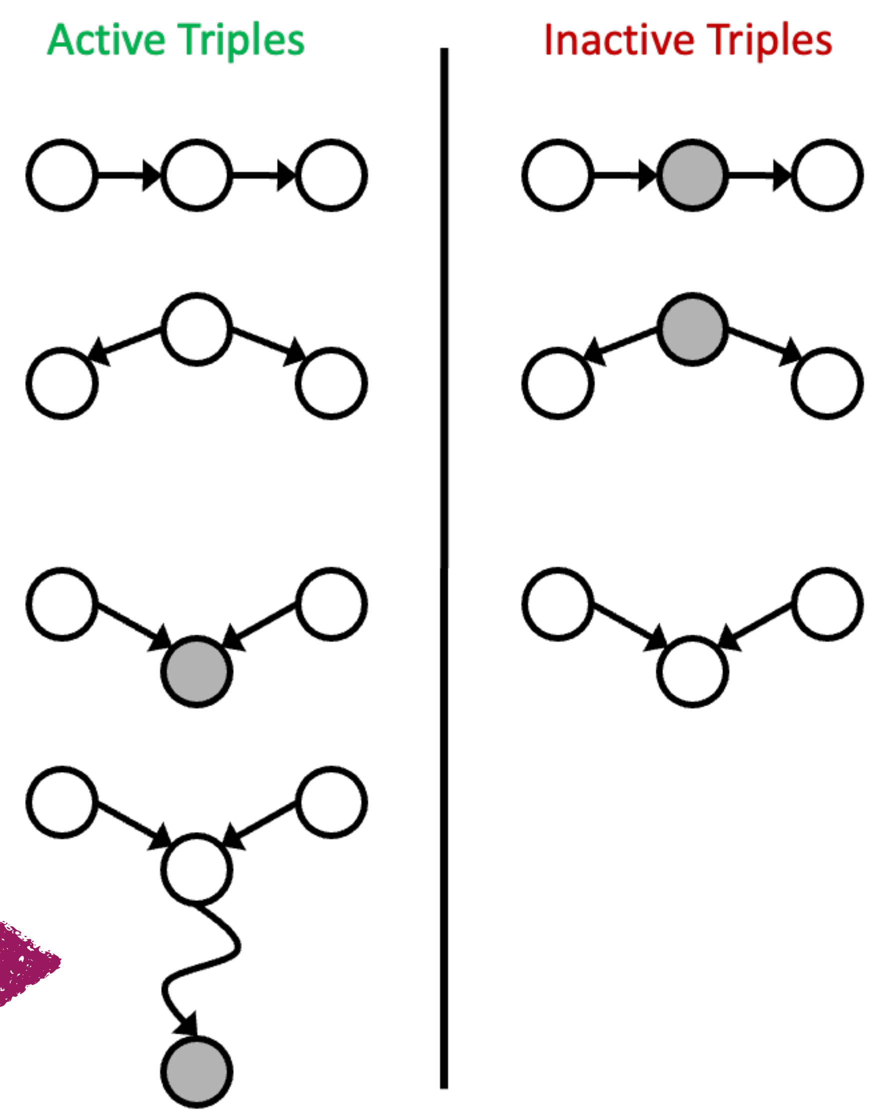
Active/Inactive Paths

- General question: Given a BN, want to know whether X and Y are conditional independent given a set of evidence {Z}.



Active/Inactive Paths

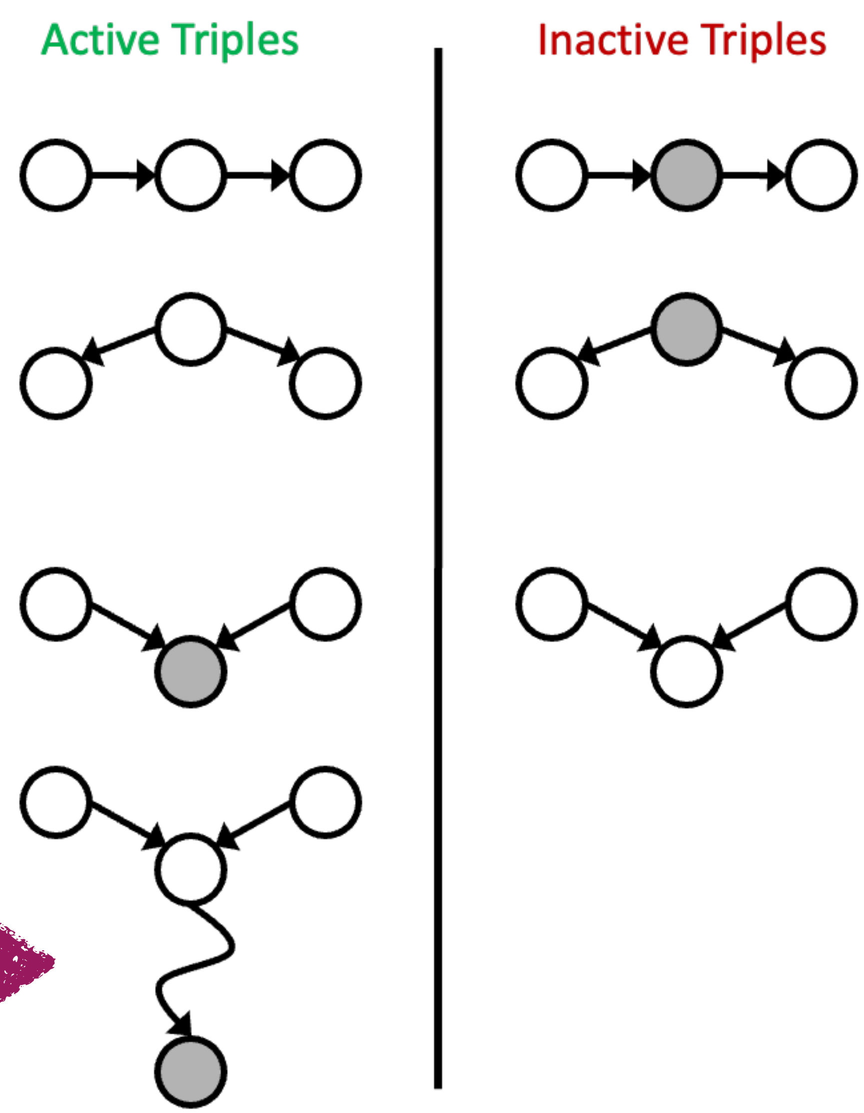
- General question: Given a BN, want to know whether X and Y are conditional independent given a set of evidence {Z}.



V-structure holds for chain of descendents!

Active/Inactive Paths

- General question: Given a BN, want to know whether X and Y are conditional independent given a set of evidence {Z}.



V-structure holds for chain of descendents!

Active path: exist a path from X to Y in which no inactive triplet exists!

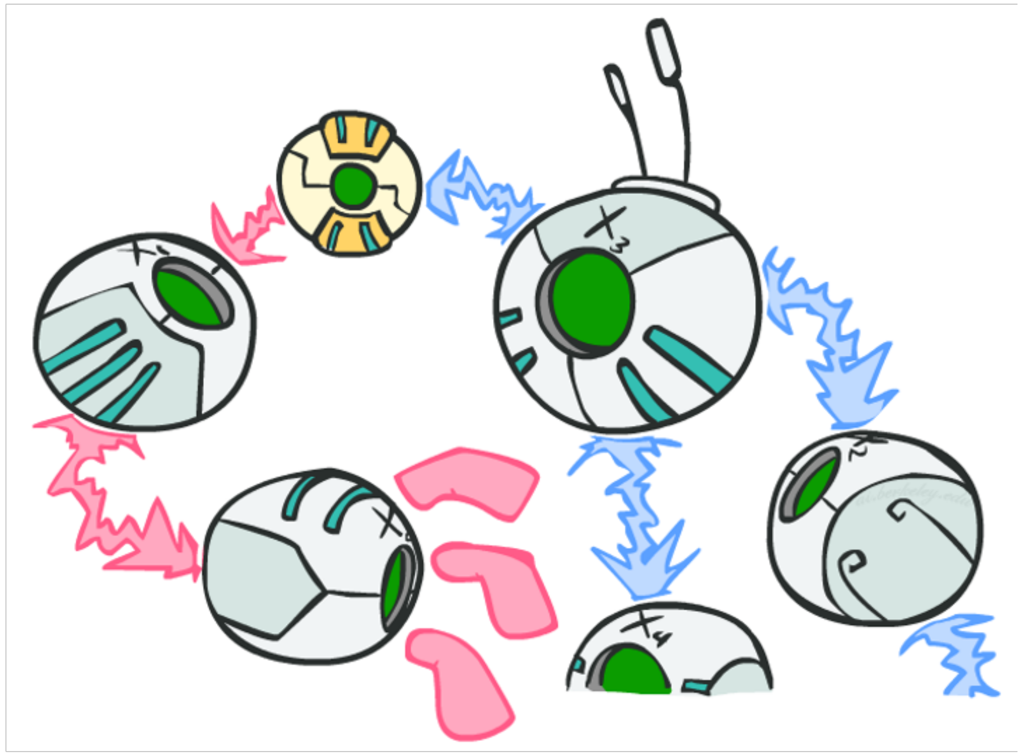
D-Separation Algorithm

- Query: $X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$
- Check all (undirected!) paths between $\boxed{X_i}$ and $\boxed{X_j}$
  - If one or more active, then independence not guaranteed

$$X_i \not\perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$$

  - Otherwise (i.e. if all paths are inactive), then independence is guaranteed

$$X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$$



Slide courtesy: Dan Klein & Pieter Abbeel

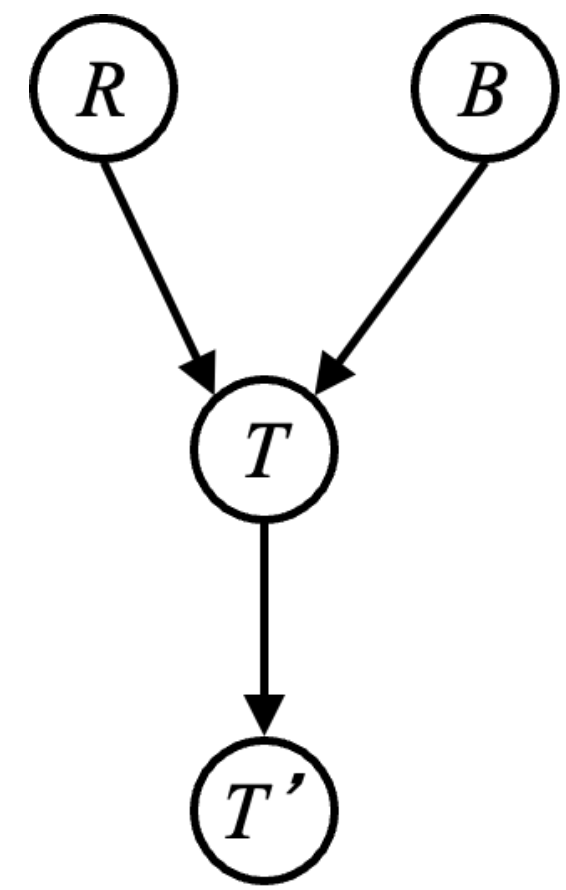
Example

- Whether the following holds?

$$R \perp\!\!\!\perp B$$

$$R \perp\!\!\!\perp B | T$$

$$R \perp\!\!\!\perp B | T'$$



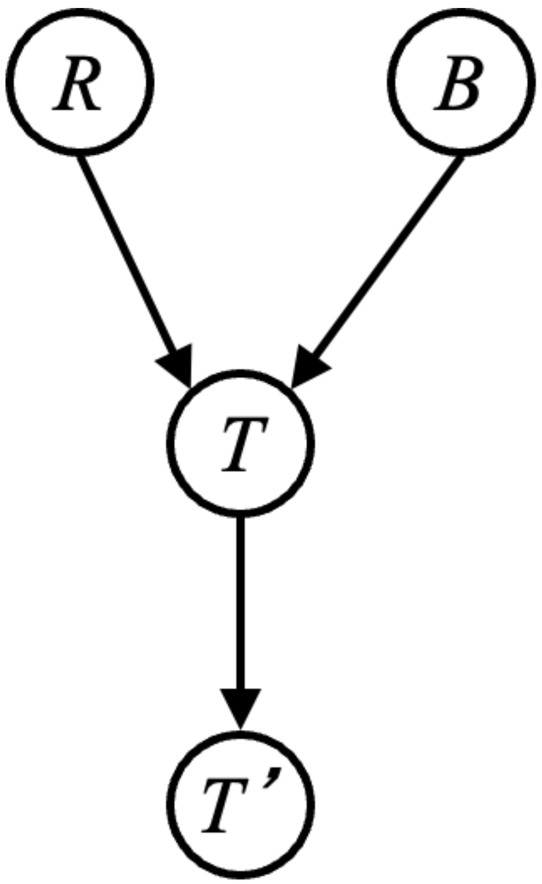
Example

- Whether the following holds?

$$R \perp\!\!\!\perp B$$

$$R \perp\!\!\!\perp B | T$$

$$R \perp\!\!\!\perp B | T'$$



Example

- Whether the following holds?

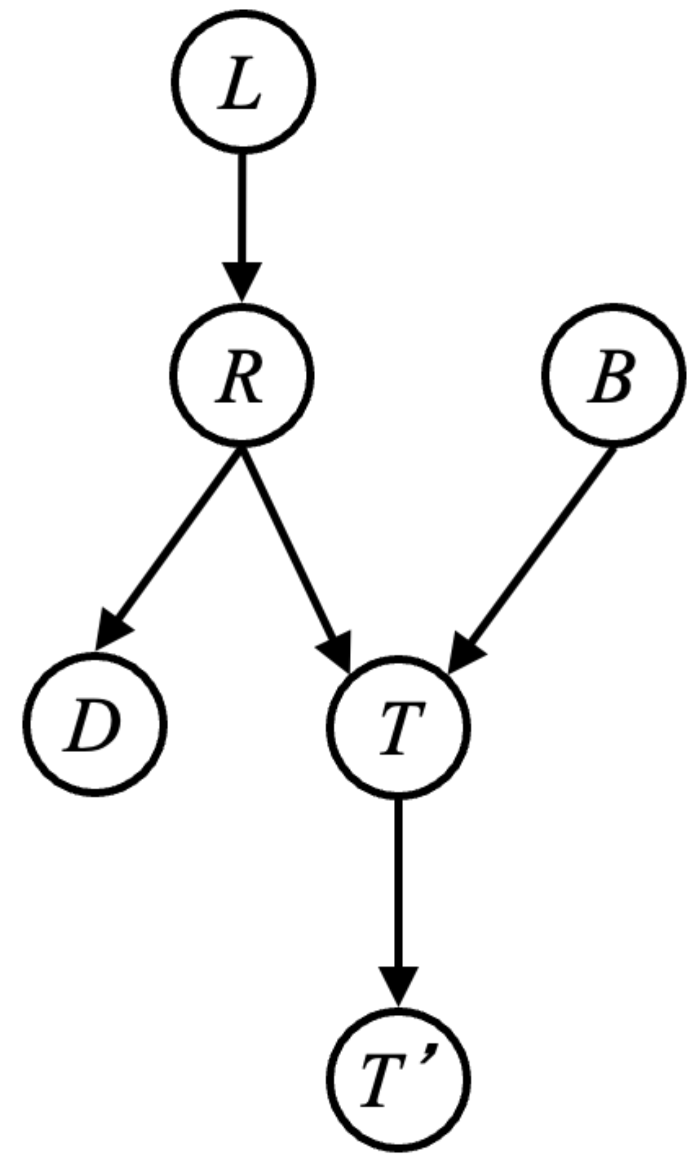
$$L \perp\!\!\!\perp T' | T$$

$$L \perp\!\!\!\perp B$$

$$L \perp\!\!\!\perp B | T$$

$$L \perp\!\!\!\perp B | T'$$

$$L \perp\!\!\!\perp B | T, R$$



Example

- Whether the following holds?

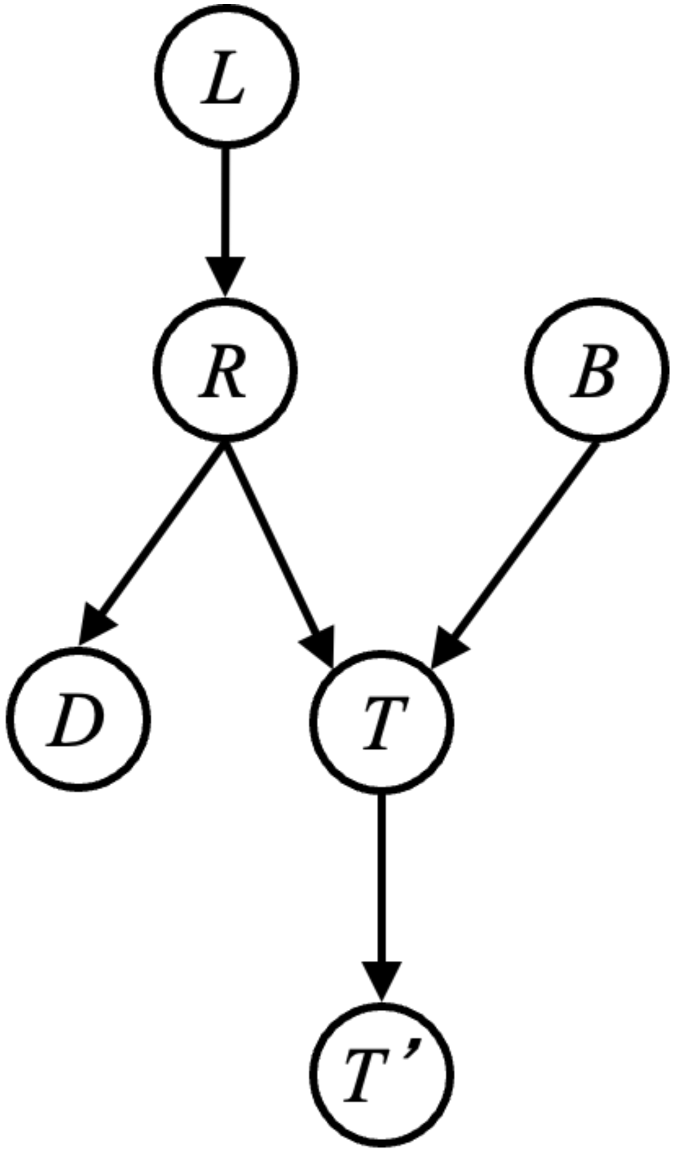
$L \perp\!\!\!\perp T' | T$ ✓

$L \perp\!\!\!\perp B$ ✓

$L \perp\!\!\!\perp B | T$

$L \perp\!\!\!\perp B | T'$

$L \perp\!\!\!\perp B | T, R$ ✓



Knowledge Reasoning: II

- Probabilistic reasoning
- The structure of probability distributions
- Bayes net: Representation
- Bayes net: Conditional independence
- Bayes net: Exact inference
- Take-home messages

Basic Tasks in Probabilistic Reasoning

- In probabilistic reasoning, we try to model the joint distribution of a set of random variables $P(X_1, X_2, ..., X_n)$ and do:
  - Inference: answering queries about the marginal distributions.
  - Conditional independence test: decide the conditional independence of a subset of random variables.
  - Learning: obtain the structure of the joint distribution.

The Inference Problem

- Inference in probabilistic models: calculate margin quantities from the joint distribution represented by the model.

- Examples:
  - Posterior probability

    $$P(Q|E_1 = e_1, \ldots E_k = e_k)$$

  - Most likely explanation:

    $$\text{argmax}_q\ P(Q = q|E_1 = e_1 \ldots)$$

Slide courtesy: Dan Klein & Pieter Abbeel

Inference by Enumeration

- General case:
  - Evidence variables: $E_1 \dots E_k = e_1 \dots e_k$
  - Query* variable: Q
  - Hidden variables: $H_1 \dots H_r$

$$\left.\begin{array}{l} E_1 \dots E_k = e_1 \dots e_k \\ Q \\ H_1 \dots H_r \end{array}\right\} X_1, X_2, \dots X_n$$

All variables

Inference by Enumeration

- General case:
  - Evidence variables: $E_1 \dots E_k = e_1 \dots e_k$
  - Query* variable: Q
  - Hidden variables: $H_1 \dots H_r$

$\left.\begin{array}{l} E_1 \dots E_k = e_1 \dots e_k \\ Q \\ H_1 \dots H_r \end{array}\right\} X_1, X_2, \dots X_n$

All variables

- We want:

** Works fine with multiple query variables, too*

$$\boxed{P(Q|e_1 \dots e_k)}$$

Inference by Enumeration

- General case:
  - Evidence variables: $E_1 \ldots E_k = e_1 \ldots e_k$
  - Query* variable: Q
  - Hidden variables: $H_1 \ldots H_r$

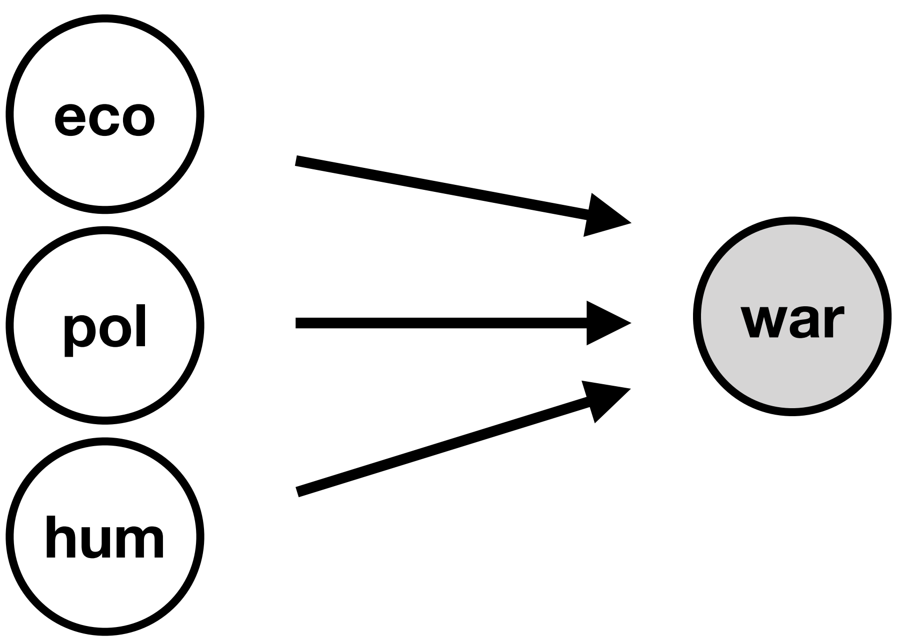
$$\left.\begin{array}{l} E_1 \ldots E_k = e_1 \ldots e_k \\ Q \\ H_1 \ldots H_r \end{array}\right\} X_1, X_2, \ldots X_n$$

All variables

- We want:

** Works fine with multiple query variables, too*

$$\boxed{P(Q|e_1 \ldots e_k)}$$

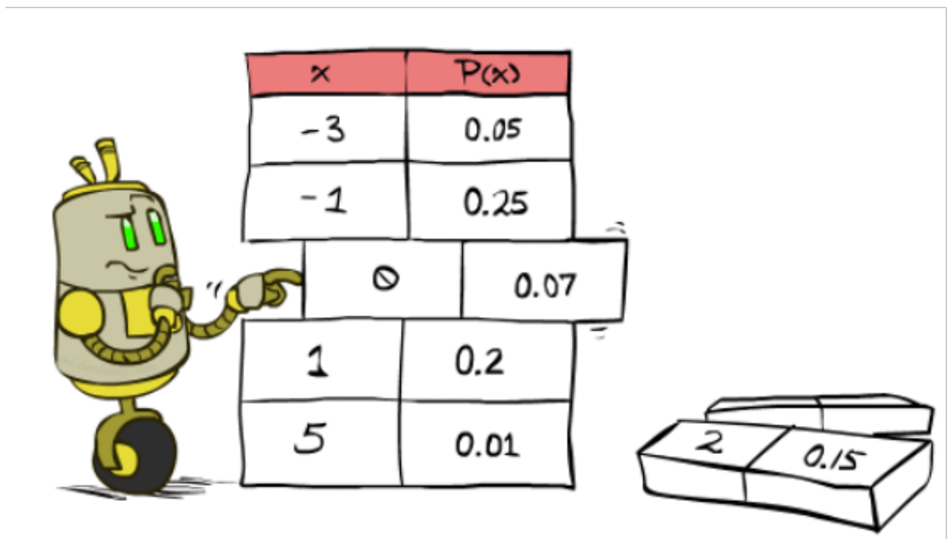


$$P(war|hum = kind)?$$

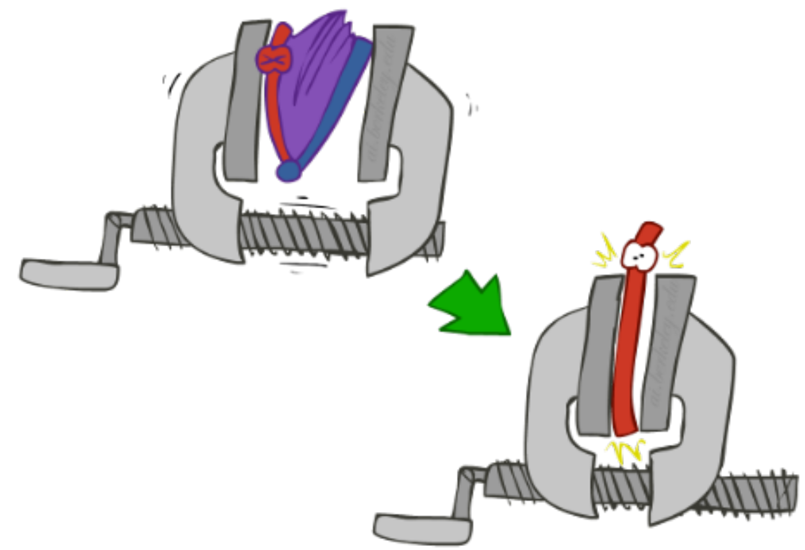
Slide courtesy: Dan Klein & Pieter Abbeel

Inference by Enumeration

- Step 1: Select the entries consistent with the evidence



- Step 2: Sum out H to get joint of Query and evidence



- Step 3: Normalize

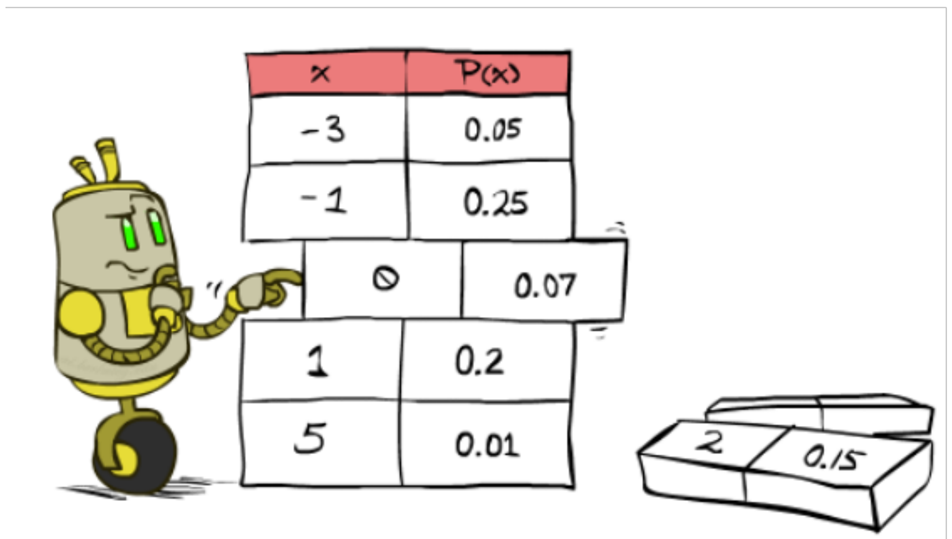
$$\times \frac{1}{Z}$$

$$Z = \sum_q P(Q, e_1 \cdots e_k)$$

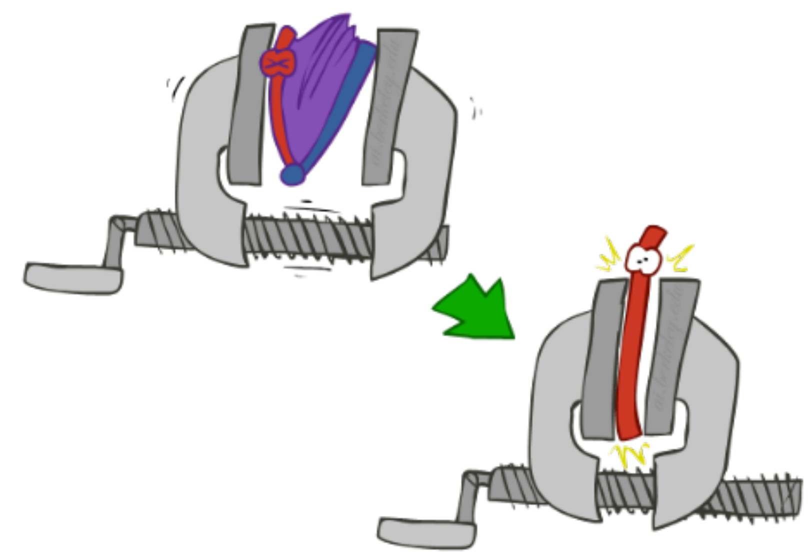
Slide courtesy: Dan Klein & Pieter Abbeel

Inference by Enumeration

- Step 1: Select the entries consistent with the evidence



- Step 2: Sum out H to get joint of Query and evidence



- Step 3: Normalize

$$\times \frac{1}{Z}$$

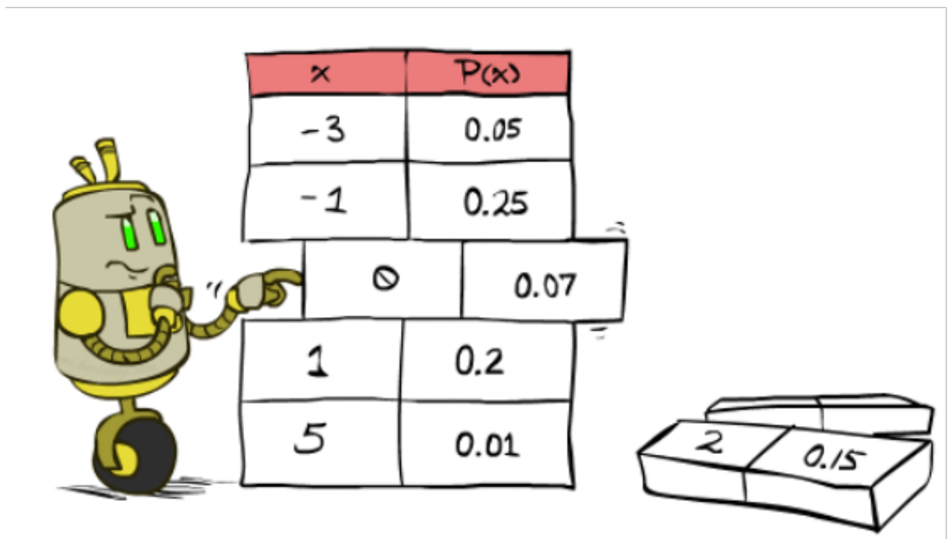
$$P(war = true | hum = kind)$$

$$= \sum_{i,j} p(war = true, eco = i, pol = j, hum = kind)/Z$$

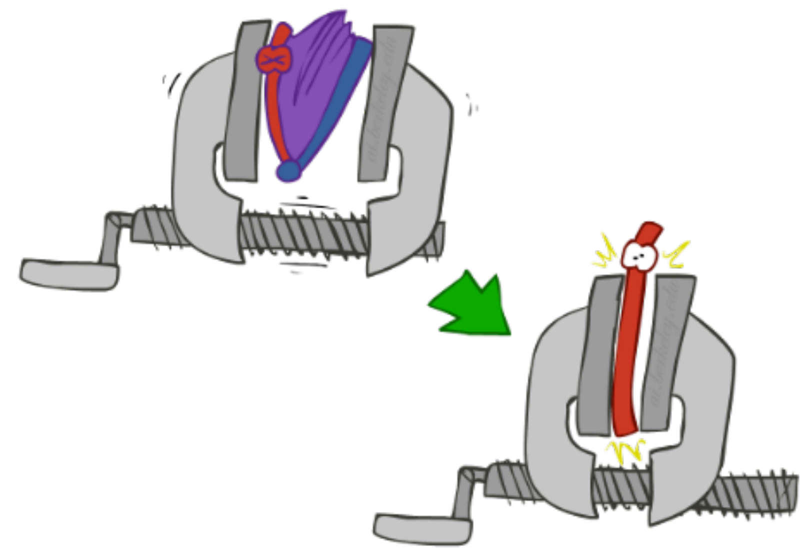
$$Z = \sum_{q} P(Q, e_1 \cdots e_k)$$

Inference by Enumeration

- Step 1: Select the entries consistent with the evidence



- Step 2: Sum out H to get joint of Query and evidence



- Step 3: Normalize

$$\times \frac{1}{Z}$$

$$Z = \sum_q P(Q, e_1 \cdots e_k)$$

$$P(war = true | hum = kind)$$

$$= \sum_{i,j} p(war = true, eco = i, pol = j, hum = kind)/Z$$

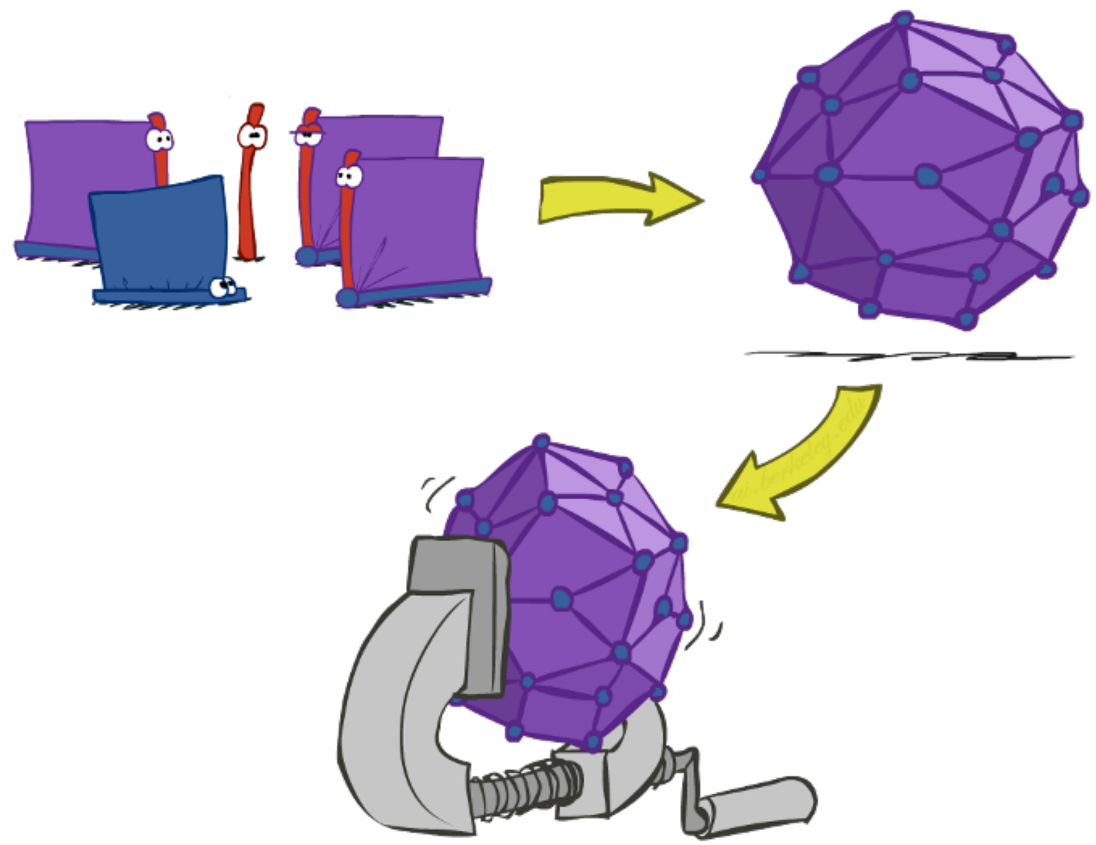
$$Z = p(hum = kind)$$

$$= \sum_{i,j,k} p(war = k, eco = i, pol = j, hum = kind)$$

Slide courtesy: Dan Klein & Pieter Abbeel

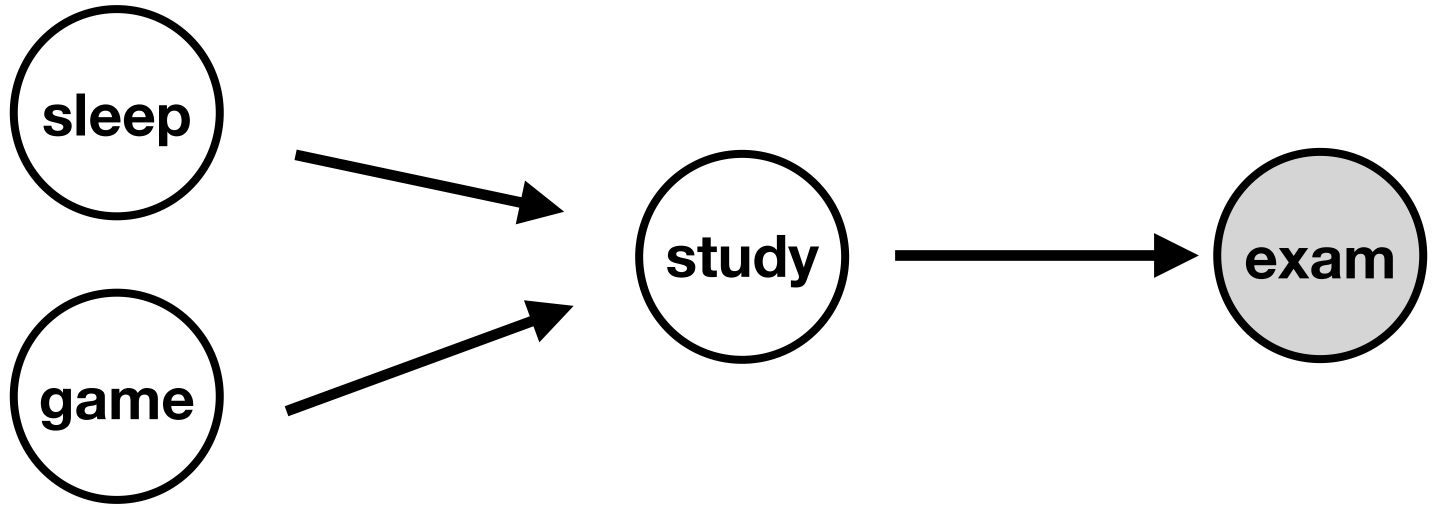
Computational Complexity

- The number of summations are exponential w.r.t. input!



We do not make use of the structure information provided by BN.

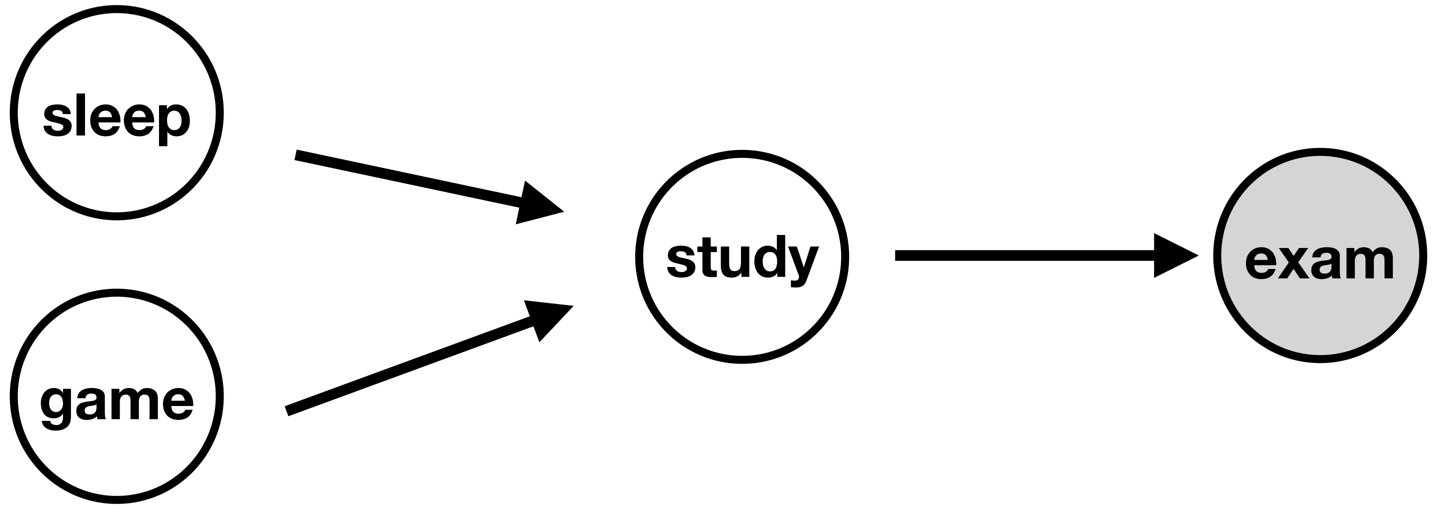
Variable Elimination



- The key idea is to marginalize early by utilizing the structure of BN.
- To answer: $p(exam|sleep = true)$
- Enumeration:

$$\sum_{i,j} p(exam|study = i)p(study = i|sleep = true, game = j)p(sleep = true)p(game = j)$$

Variable Elimination



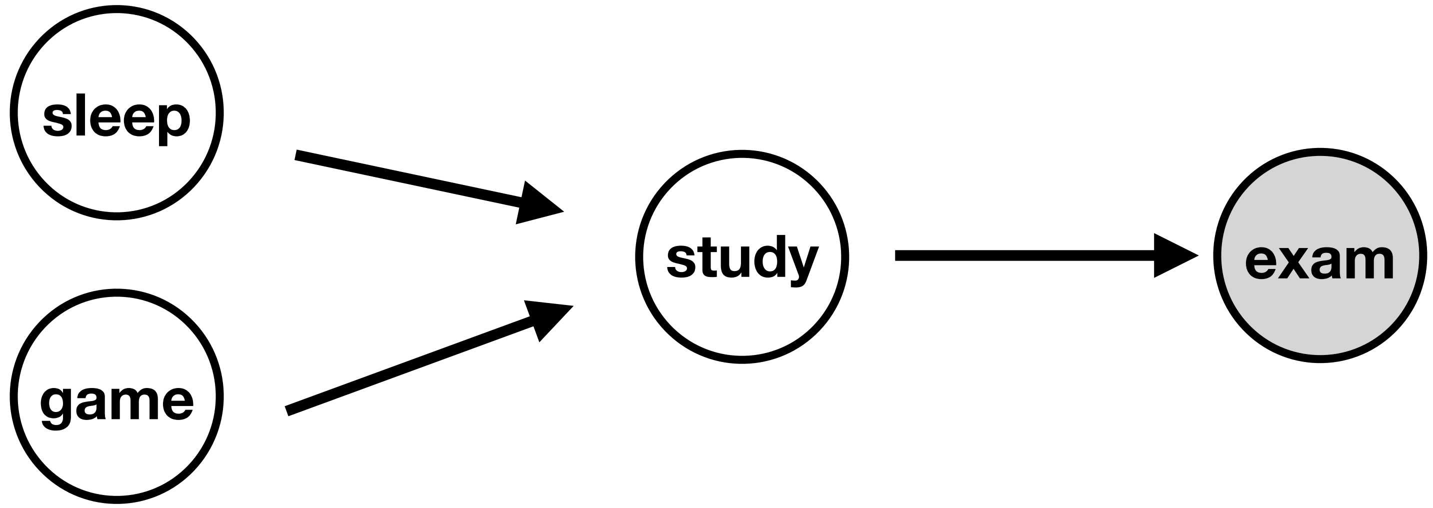
- The key idea is to marginalize early by utilizing the structure of BN.
- To answer: $p(exam|sleep = true)$
- Enumeration:

$$\sum_{i,j} p(exam|study = i)p(study = i|sleep = true, game = j)p(sleep = true)p(game = j)$$

- Elimination:

$$\sum_{i} p(sleep = true)p(exam|study = i) \sum_{j} p(study = i|sleep = true, game = j)p(game = j)$$

Variable Elimination



- The key idea is to marginalize early by utilizing the structure of BN.

- To answer: $p(exam|sleep = true)$

- Enumeration:

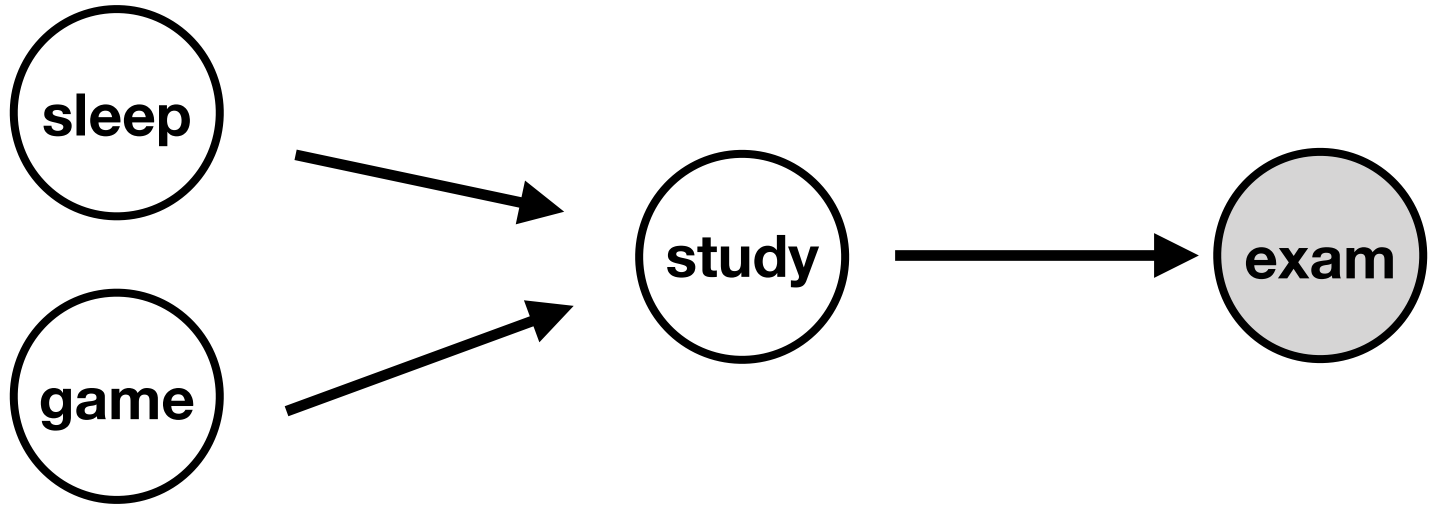
$$\sum_{i,j} p(exam|study = i)p(study = i|sleep = true, game = j)p(sleep = true)p(game = j)$$

- Elimination:

Sum out hidden variables early!

$$\sum_{i} p(sleep = true)p(exam|study = i) \sum_{j} p(study = i|sleep = true, game = j)p(game = j)$$

Variable Elimination



- The key idea is to marginalize early by utilizing the structure of BN.
- To answer: $p(exam|sleep = true)$
- Enumeration:

$$\sum_{i,j} p(exam|study = i)p(study = i|sleep = true, game = j)p(sleep = true)p(game = j)$$

- Elimination:

Sum out hidden variables early!

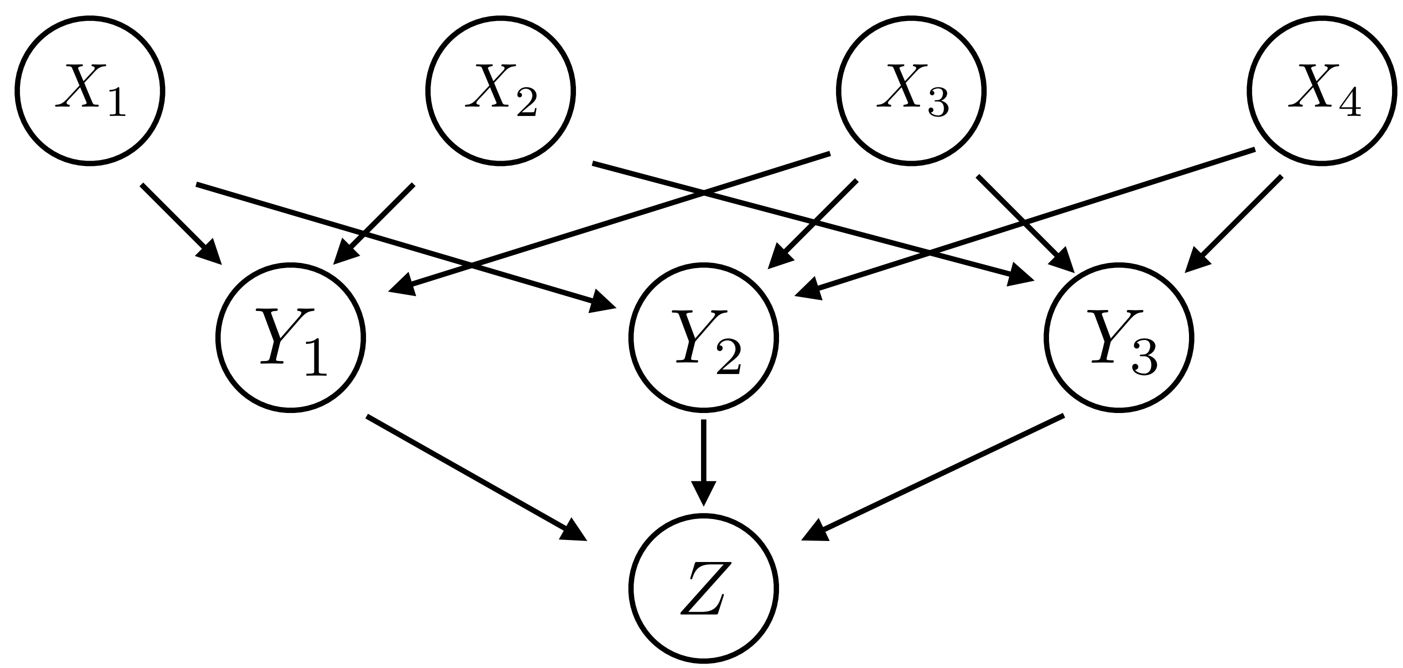
$$\sum_{i} p(sleep = true)p(exam|study = i) \sum_{j} p(study = i|sleep = true, game = j)p(game = j)$$

Can save much by avoiding later enumeration!

Computational Complexity

- Is variable elimination a poly-time algorithm?
- Consider representing 3-CNF as BN:

$$(X_1 \vee X_2 \vee \vee \neg X_3) \wedge (\neg X_1 \vee X_3 \vee X_4) \wedge (X_2 \vee \neg X_3 \vee X_4)$$



If we can determine whether $p(Z) = 0$, then we can solve the 3-SAT problem which is NP-complete. So the exact inference in BN is NP-hard.

Knowledge Reasoning: II

- Probabilistic reasoning
- The structure of probability distributions
- Bayes net: Representation
- Bayes net: Conditional independence
- Bayes net: Exact inference
- Take-home messages

Take-Home Messages

- Probabilistic reasoning models the real world with a joint probability distribution of random variables.
- Bayes nets are acyclic directed graphs representing the factored joint probability distributions.
- The conditional independence properties of BNs can be discovered by D-separation structures.
- The variable elimination algorithm makes use of the factored structure represented by BN to do more efficient exact inference: marginalize early. But exact inference is always NP-hard.

Next lecture: approximate inference in BN & causal inference

Thanks for your attention!
Discussions?

Acknowledgement: Many materials in this lecture are taken from
http://ai.berkeley.edu/lecture_slides.html